DOCUMENT RESUME

ED 428 075 TM 029 464

AUTHOR	Glass, Gene V., Ed.
TITLE	Education Policy Analysis Archives, Volume 5, 1997.
ISSN	ISSN-1068-2341
PUB DATE	1997-12-00
NOTE	673p.; "Education Policy Analysis Archives" is an electronic-only journal (see http://olam.ed.asu.edu/spaa) covered on an article-by-article basis in "Current Index to Journals in Education" (CIJE). For the 22 articles in this volume, see TM 521 583-604 (in CIJE).
AVAILABLE FROM	Web site: http://olam.ed.asu.edu/epaa/vol5.html
PUB TYPE	Collected Works - Serials (022)
JOURNAL CIT	Education Policy Analysis Archives v5 n1-22 1997;
EDRS PRICE	MF04/PC27 Plus Postage.
DESCRIPTORS	Educational Change; *Educational Policy; Educational Research; Electronic Journals; Elementary Secondary Education; Foreign Countries; Higher Education; *Research Methodology; Scholarly Journals; Teacher Evaluation; Teaching Methods

ABSTRACT

This electronic journal covers a variety of issues related to educational policy. The 22 articles (issues) of Volume 5 analyze aspects of policy in the United States and other countries for elementary, secondary, and higher education. Articles include: (1) "Markets and Myths: Autonomy in Public and Private schools" (Glass); (2) "Where Have All the Teachers Gone?" (Fetler); (3) "Testing Writing on Computers" (Russell and Haney); (4) "Service-Learning in Costa Rica and Indonesia" (Williams and Eiserman); (5) "'Out of Our Minds':A Review" (Clifford); (6) "Academic Freedom, Tenure, and Student Evaluation of Faculty" (Haskell) [and commentary]; (7) "The Circle of Learning: Individual and Group Processes" (Chang and Simpson); (8) "Response to Haskell" (Stake); (9) "The History of the Reserve Officer Training Corps among the Association of American Universities from 1982 to 1992: Review of Institutional Responses to ROTC Policy Regarding Homosexuals" (Deumer); (10) "Cultural Differences and the Construction of Meaning" (Pena); (11) "Centralized Goal Formation and Systemic Reform" (Strike); (12) "Studying the Rural in Education" (Howley); (13) "'Qualitative Research Methods': An Essay Review" (McLean, Myers, Smilie, Vaillancourt); (14) "Rhetoric and Educational Policies and the Use of History for Citizenship Education" (Wong); (15) "Early Childhood Evaluation and Policy Analysis: A Communicative Framework" (Wallat and Piazza); (16) "Staffing Up and Dropping Out" (Felter); (17) "Academic Freedom, Promotion, Reappointment Tenure and The Administrative Use of Student Evaluation of Faculty (Part II)" (Haskell); (18) "Academic Freedom, Promotion, Reappointment, Tenure and The Administrative Use of Student Evaluation of Faculty (SEF): Part III" (Haskell); (19) "Comprehensive Study of Factors Impacting Perceived Quality in School Organizations" (Poston), Jr.; (20) "The Use and Abuse of Socrates in Present Day Teaching" (Rud, Jr.) [and Commentary]; (21) "Academic Freedom, Promotion, Reappointment, Tenure and The Administrative Use of Student Evaluation of Faculty (SEF): Part IV" (Haskell); (22) "Changing Definitions and Off-Loading Responsibility on Alberta's Post-Secondary System" (Barnetson). (SLD)

VOLUME 5

ISSN 1068-2341

ED 428 075

Education Policy Analysis Archives

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
● Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Gene Glass
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
1

editor:
Gene V. Glass
Arizona State University

TM029464

BEST COPY AVAILABLE

VOLUME 5 -
ISSN 1068-2341

Education Policy Analysis Archives

editor:
Gene V. Glass
Arizona State University

TM029464

EPAA • enter the archives • browse the abstracts • the editors • the edit board
• submit article • submit commentary • search • subscribe
volume: • 1 • 2 • 3 • 4 • 5 • 6 • 7

EPAA Editors

EPAA is Edited by Gene V Glass, who is the Associate Dean for Research in the College of Education at Arizona State University. You may reach the Editor in a variety of ways:

- Email to glass@asu.edu
- Paper mail to College of Education, Arizona State University, Tempe AZ 85287-0211
- Voice messages are taken at 602-965-9644

The *EPAA* **Associate Editor for Spanish Language** is Roberto Rodríguez Gómez of the Universidad Nacional Autónoma de México. His email address is roberto@servidor.unam.mx. Articles in Spanish may be submitted either to Editor Glass or Associate Editor Rodríguez. Associate Editor Rodríguez will supervise the review and make final editorial decisions on Spanish language articles.

The **Book Review Editor** for *EPAA* is Walter E. Shepherd. He can be reached at shepherd@asu.edu

The **Commentary Editor** for *EPAA* is Casey D. Cobb, an Assistant Professor in the Department of Education at the University of New Hampshire. He can be reached at casey.cobb@unh.edu

We thank Charalambos Vrasidas (pambos@asu.edu) for his considerable expertise in graphic design for giving this site its appealing appearance. Kyle Rosenthal assisted at many points in the development of this site with consultation on systems design and administration.

archives | abstracts | editors | board | submit | comment | subscribe | search

Volume 5 1997

EDUCATION POLICY ANALYSIS ARCHIVES

Article Pages

1 Glass: Markets & Myths: Autonomy in Public & Private Schools 38
2 Fetler: Where Have All the Teachers Gone? 17
3 Russell & Haney: Testing Writing on Computers 20
4 Williams & Eiserman: Service-Learning in Costa Rica and Indonesia 30
5 Clifford: Review of Howley, Howley & Pendarvis Out of Our Minds 6
6 Haskell: Academic Freedom, Tenure, and Student Evaluation of Faculty 32
Commentary 17
7 Chang & Simpson: The Circle of Learning: Individual and Group Processes 20
8 Stake: Response to Haskell's "Academic Freedom ... & Student Evaluation" 6
Commentary 21
9 Duemer: ROTC Policy Regarding Homosexuals 19
10 Peña: Cultural Differences and the Construction of Meaning 18
11 Strike: Centralized Goal Formation and Systemic Reform 42
12 Howley: Studying the Rural in Education 13
13 McLean et al.: Qualitative Research Methods: An essay review 38
14 Wong: Rhetoric and Educational Policies on the Use of History 20
15 Wallat & Piazza: Evaluation and Policy Analysis: A Communicative Framework 47
16 Fetler: Staffing Up and Dropping Out 23
17 Haskell: Student Evaluation of Faculty: View from the Court 47
18 Haskell: Student Evaluation of Faculty: Psychometric Validity of Court's Views 51
19 Poston: Factors Impacting Perceived Quality in School Organizations 49
20 Rud: Use & Abuse of Socrates in Teaching 16
Commentary 4
21 Haskell: Administrative Use of Student Evaluation of Faculty 42
22 Barnetson: Off-loading Responsibility in Alberta's Post-Secondary System 18

EPAA
• enter the archives • browse the abstracts • the editors • the edit board
• submit article • submit commentary • search • subscribe
volume: • 1 • 2 • 3 • 4 • 5 • 6 • 7

This article has been retrieved 5977 times since January 6, 1997.

Education Policy Analysis Archives

Volume 5 Number 1 **January 6, 1997** **ISSN 1068-2341**

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass Glass@ASU.EDU.

College of Education

Arizona State University,Tempe AZ 85287-2411

Copyright 1997, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

Markets and Myths: Autonomy in Public and Private Schools

Sandra Rubin Glass
Arizona State University

Abstract

School choice is the most controversial education policy issue of the 1990s. John Chubb and Terry Moe's *Politics, Markets and America's Schools* stimulated this investigation. They concluded that teacher and administrator autonomy was the most important influence on student achievement. They assumed that the organization of private schools offered greater autonomy resulting in higher student achievement and that the bureaucracy of public schools stifles autonomy limiting student achievement. The research undertaken here elaborates, elucidates, and fills in the framework of teacher and principal autonomy in public and private secondary schools. Interviews of more than thirty teachers and administrators in six high schools, observations, field notes, and analysis of documents collected in the field form the empirical base of this work. The sites included three private, independent, nondenominational secondary schools which are college preparatory and three public secondary schools noted for high graduation rates and offering numerous advanced placement courses.

The feelings expressed by both public and private school participants in this study testify to equally high degrees of autonomy. Issues that emerged from data analysis in this study which mitigate and shape autonomy include the following: conflicting and contradictory demands, shared beliefs, layers of protection a system of laws, funding constraints and matters of size of the institution. These issues challenge oversimplified assertions that differences of any importance exist between the autonomy experienced by professionals in public and private high schools. This study reveals the complexity of the concept of autonomy and challenges the myth

that teachers and principals in private schools enjoy autonomy and freedom from democratic bureaucracy that their public school counterparts do not.

Table of Contents

- The Problem
- Review of Literature
- Methods of the Study
- The Schools in the Study
- Findings
- Summary & Conclusions
- References
- The Interviews

About the Author

Sandra Rubin Glass

sandy.glass@asu.edu

Sandra Rubin Glass is a Faculty Associate in the College of Education at Arizona State University where she received her PhD in Educational Leadership & Policy Studies in 1993. Her M.A. (1974) in the Teaching of Earth Sciences is from Northeastern Illinois University; her Bachelors degree (1968) is from National-Louis University (formerly National College of Education).

Dr. Glass holds an appointment in the Office of Professional Field Experiences at ASU where she develops and implements training and workshops for student and beginning teachers. She is a former teacher and administrator in both public and private schools, and once taught on the Navajo Indian Reservation.

Copyright 1997 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://olam.ed.asu.edu/epaa

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey@olam.ed.asu.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Marshall University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Rocky Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
Arizona State University

Dennis Sayers
University of California at Davis

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina--Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher Education

Mary P. McKeown
Arizona Board of Regents

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois--UC

Robert T. Stout
Arizona State University

---

archives | abstracts | editors | board | submit | comment | subscribe | search

Background

School choice, the prerogative of parents to elect the school their children attend, is the most controversial education policy issue of the 1990s. As many as ten states had formally adopted open enrollment or educational choice provisions by Spring 1991. In the general election of 1992, Colorado voters resoundingly voted down a proposed voucher amendment to the state constitution that would have permitted choice between public and private schools. Several other states currently offer informal open enrollment or choice provisions or are planning legislation that would mandate some form of school choice (Bierlein, Sheane & McCarthy, 1991). The publication in 1990 of Chubb and Moe's *Politics, Markets & America's Schools* gave pro-choice advocates a rallying point. Chubb and Moe argued that sense of autonomy and freedom from bureaucratic pressure are the most powerful determinants of a school's success in advancing academic learning. They asserted that these conditions are more prevalent in private schools (which "seem to be better performers," p. 24) than in public schools because of "market forces" (p. 37). Although their analysis looked at the relationship of autonomy and student achievement *only within public schools* (because of what they regarded as limitations of their data sources, namely the High School and Beyond Survey), Chubb and Moe argued that teacher and principal autonomy should be greatly different between public and private sector schools. They advanced no evidence on this question of public and private school differences, but merely speculated about it. Although Chubb and Moe confined their final recommendations to the public sector (acknowledging that privatizing America's schools is an impractical fantasy), their work has become the foundation for many who would extend school choice to encompass both public and private schools.

School choice advocates believe that educators in public schools are constrained by bureaucracy from acting to improve the conditions of education. They assert that, in contrast, principals and teachers in private schools enjoy greater autonomy and have more control over significant decisions like teaching methods, curriculum and personnel. Consequently, private schools produce a higher quality educational experience for the students. They view private school educators as responding primarily to the needs of pupils and the directly expressed wishes of parents: in effect, responding to their clients' expressions of market preferences. They view public school educators as locked into a bureaucracy that stifles initiative and effort and insulates employees from public pressures. This view of the difference between public and private education may be greatly oversimplified. Private school teachers may respond to parents, whereas public school teachers may respond to parents acting as a "school board." The difference may be important, or it may be a distinction without an important difference. Moreover, the needs created by children's circumstances may motivate public school teachers as strongly as they motivate teachers in private schools. To the extent to which teachers in private and public schools encounter children with similar needs, they may act similarly. What seems to have occurred in thinking about public and private education is what Freire described as "mythicizing reality": attempting to conceal certain facts which explain the way persons exist in the world (1981, p. 22). That the myths have been successfully transmitted is evidenced by journalistic excesses such as the following: "Study after study has shown that the more schools are freed of outside bureaucratic control, the more likely they are to succeed." (editorial, "The Arizona Republic," November 12, 1991)

It may be that the size and complexity of an organization, like a school, determines the degree of autonomy felt by participants rather than whether its governance is private or public. Moreover, in today's complex world, public controls of many and varied types often extend to private institutions, because they participate through loan programs or grants or contracts in governmental programs or fall under the jurisdiction of courts. Chubb and Moe (1990), Coleman and Hoffer (1987), and most others attempting to think about the implications of organization and control for administrator and teacher actions are merely engaging in speculation about that which few have studied directly.

The Problem of Teacher and Principal Autonomy

The study reported here did not assume that there are clear and distinct differences between workplace conditions in public and private schools. Rather, it attempted to bring to the surface those conditions which otherwise may be overlooked or neglected, conditions which constrain teacher and principal autonomy in both private and public schools. Another aspect of the problem under investigation here is to detail the differences that may or may not exist between principals' and teachers' sense of autonomy and control in private and public school settings. The need for such an investigation is clear to persons on all sides of the issue of school choice: "Case studies and other more narrowly focused research into schools could help develop an understanding of these relationships that could guide both educators and policymakers in determining the appropriate role of autonomy in school improvement" (Glass & Matthews, 1991, p. 26).

To limit the scope of this study, a specific type of private school was selected from the braod range of possibilities. The college preparatory independent school without denominational affiliations was chosen because it operates in an environment more like the legal description of a public school located in an upper-middle class community. The student populations are similar in many social and economic circumstances, both having students whose parents reflect the behaviors of those with the greatest choice of school type.

Return to the Table of Contents

What Has Already Been Written on this Problem

The literature on bureaucracy and autonomy is huge within sociology, and substantial within the specialty of the sociology of education. The works of Powell (1990), Lightfoot (1983), Ball (1987), Sedlak and others (1986), McNeil (1986), and Firestone and Bader (1991) are most influential. This literature, at times tenuously connected to empirical research, holds ideas and conceptions of teacher autonomy that may foreshadow many of the ideas to be encountered in the field in the pursuit of this research.

Private schools appear to be subject to fewer apparent constraints than those encountered by public schools. Their governance and financing make them directly responsible to a constituency which they must satisfy to stay solvent (Powell, 1990). They have no direct obligation to the whole of society (Grant, 1988).

In contrast, public schools must serve the needs of children as seen by elected or appointed representatives of the public at local, state and federal levels. While such external government mandates, court decisions, and union contracts have, perhaps, a marginal impact on the independent school, certainly these schools are not immune from public regulations concerning health, safety, and civil rights. Nor are they protected from the not-so-subtle intrusions of publishers, external testing, and especially college admissions requirements (Powell, 1990). The subtle curricular power of the Advanced Placement (AP) examinations, for example, exerts a pressure on the curriculum and accountability of private high school teachers similar to that of state mandated testing in the public sector (Powell, 1990).

Surely both public and private schools are subject to organizational constraints that stem from "external structures (subjects, periods of time) . . . occupational norms (order in the classroom, class rules and so on) . . . [that ensure] some minimal level of uniformity" (Elmore, 1987, p. 64). Ball (1987) went further and suggested that educators ask, "How autonomous is the organization and its actors from its clients, publics, superiors and audiences or the basic social and economic structures of the society?" He suggested the notion of relative autonomy: " . . . organizations are not independent or self-sufficient phenomena" (p. 247).

There is a generally held perception by society that private schools are successful educational businesses. This is not necessarily the case. In preparation for her research on public and private high schools, Lightfoot (1983) participated in a scholars' seminar. Invited scholars included those whose work centered on the history, policies, and practices of schools. The assumption that all private schools are thriving was called into question.

> But the common assumption that the private schools were thriving and flourishing was unsettling, and was experienced by some members [of the seminar] as a disregard for the great variation in success and resources among them (p.8).

Indeed, Chubb and Moe (1985) concluded, "Relative to public schools, private schools appear to delegate significant discretion to their teachers, and to involve them sufficiently in school level policy decisions to make them feel efficacious" (p. 37). This common sense mythology perpetuates the misconception that the private school community shares one view of what constitutes a good school. The reality may be, as Powell (1990) suggested, that private schools often vary sharply in content and process and espouse a wide variety of purposes (single-sex schools, boarding schools, schools that cater to a particular ability level) based on the type of community they serve. While public schools are traditionally depicted as being more diverse, this diversity is more a matter of economic differences. There are schools for the poor, schools for the middle class, and schools for the suburbanites. Within the private school context of general affluence there is more cohesiveness of purpose, and more shared experience. Image and reliance on this mythology of academic excellence may be what allows some private schools to compete in the marketplace along with other private and public schools (Powell, 1990).

Autonomy may be experienced by teachers as a school runs smoothly and little administrative attention is apparent. While teachers may enjoy a great deal of autonomy in these circumstances, their autonomy may actually operate within a narrow range of discretion (McNeil, 1984). The degree of discretion may rest with the administrator. In addition, Corbett found that, "Community preferences lurk constantly at the borders of the school organization, and the superintendent and the principal are the entry points" (Corbett, 1991, p. 93). Indeed, Chubb and Moe's (1990) enthusiasm for autonomy results from their discovery of a statistical correlation between "autonomy" and "student achievement test score gains." They assumed that the causal influence ran from the former to the latter. Glass and Matthews (1991) contended that it was even more likely that the causal direction was reversed in that teachers and principals were granted more autonomy when their test scores were in good shape. Hence, they suggested that it may be achievement levels causing autonomy to be granted rather than the other way around. If Glass and Matthews are correct and Chubb and Moe are not, then granting autonomy would not be expected to result in increased achievement, nor would more autonomous private schools enjoy *ipso facto* greater effectiveness.

Indeed, the role played by the administrator is a key element in teacher autonomy or the reform initiative of "empowering" teachers (Powell, 1990; Lightfoot, 1983; McNeil, 1984). Powell pointed out that it is not clear how empowered teachers can coexist with strong site-based managers, a primary requirement of a private school head (p. 130). Apple and Teitelbaum (1986), however, found that within Weick's model of a loosely-coupled organization different types of professional can retain control and authority without changing or being changed by the decisions of other professionals. Teachers in any school organization are free to conduct their individual classrooms as they see fit without reducing the autonomy of the principal.

Although private school teachers may be freer of distant bureaucratic rules, regulations and procedures, they are subject to the pervasive authority of a headmaster and school board of directors. Based on the wide discrepancy between the salary of the administrator and teachers in private schools, the power exercised by the headmaster is considerable, perhaps even greater than that held by the public school principal. In some schools, observed Lightfoot (1983), the "unquestionable dominance and benign power" of the head only underscores the faculty's "relative powerlessness and reinforces the childlike impulses" (p. 341). Even in the case of more democratic and benign leaders, private school teachers are well aware that reappointment and references for one's resume depend on satisfying the head (Baird, 1977). Since services must be "sold to potential clients," some teachers may find themselves "caught between incompatible interpretations of their own self-interest" (Ball, 1987, p. 269).

It may not be possible to understand teacher autonomy merely from examining the obvious governmental or organizational forms that are set up to direct their actions. The working conditions of both public school and private school teachers may contain any number of what appear to be constraints on their autonomy: federal, state, and district policy; school board and administrator demands; pressures from state mandated or college testing (Noble and Smith, 1994; Smith and Rottenberg, 1991) ; the need to please parents, students, other teachers, and community; standardization practices such as career ladders (Firestone and Bader,1991; Popkewitz and Lind, 1988). But how teachers manage those constraints is crucial in defining their work life. Sedlak and others (1986) pointed out that, historically, teachers acquiesce to centralized authority yet, once they close their classroom door, most teachers are able to exercise enormous discretion. The current spate of reform initiatives produces constraints which treat teachers as passive receivers of external advice and undermine their professional authority. Elmore suggested that rather than reform, the "result is teacher resistance and student disengagement" (1987, p. 60). Faced with challenges to their autonomy, some imaginative teachers "have used their ingenuity and skill in order to arrive at a way out" (Kozol, 1981, p. 51) or participated in the "strategy of 'omissive action' (like non-cooperation . . .)" (Ball, 1987, p. 268). Indeed, Feiman-Nemser and Floden asserted that, based on their review of several studies of teacher culture, current research replaces the image of "a passive teacher molded by bureaucracy and buffeted by external forces" with the image of "an active agent, constructing perspectives and choosing actions," (1986, p. 523).

Return to the Table of Contents

Methods

The methods employed in this investigation were those of the multi-site qualitative case study: interviews from multiple data sources, observations and field notes from a variety of on-site meetings and visits, and analysis of documents (brochures, teacher handbooks, policy manuals, meeting agendas). An intensive study of three secondary schools of each type (public and private) was conducted. Fourteen private school teachers, fifteen public school teachers and their associated principals, heads, and assistants were interviewed at each site.

Site Selection

To sharpen the boundaries of this exploration, I focused on the type of private schools known as "independent" schools. There are two analytic advantages to this selection (Powell, 1990, p. 113). First, compared to the full range of private schools, independent schools are less inculcated with denominational religion and, therefore, operate in an environment more like the legal circumstances of public schools. Because they are the most expensive of the private schools, they are chosen by families who can afford any type of school, public or private. The fact that such schools serve primarily high-income families reflects a population that has the financial ability to choose a type of school based on preference provides the second advantage. They are the most privileged private schools, which then served as a guide by which to identify particular public schools which became their comparison group. Public schools located in the most affluent school districts serve a type of high income family similar to those who patronize independent schools. These public schools, therefore, came under consideration for selection for this study.

Another criterion for site selection was secondary schools that focus on academics. The selection of private schools was based on their accreditation or application for accreditation by North Central Association as college preparatory (North Central Association of Colleges and Schools, 1990). A focus on academic excellence is the key marketing strategy of these schools as evidenced in brochures describing their mission to prospective clients. Graduation rates and percentages of students applying to college are the best descriptors of academic excellence in the public high schools. The state department of education and the district office of each of the three public secondary schools served as sources for these data. Recommendations by educational specialists were also given consideration.

The three private schools selected were located in two cities in the same state in the southwestern United States. A similar process was used to gain access to public secondary schools; however, the initial contact was sometimes made through a district level administrator. The three public schools are located in very different sections of the same city and vary in the length of time they have been in existence.

Informant Selection

Interview data were collected from approximately thirty teachers and their associated administrators at each school. The number of informants contacted was guided by the understanding that each participant provides information about the conditions under which they and their colleagues work; therefore, they actually inform the researcher of the actions and beliefs of a few or many more.

The line which separates administrator from teacher in the private school is often unclear since many serve in both capacities. In one school selected, it is required that each administrator, except for the head, teach at least one class. It was explained that this requirement kept all administrators in touch with the needs of the students. Coincidentally, the requirement has some financial advantages for the school as well. In these cases, those who were labeled teachers are those whose primary responsibility is teaching. If only one class was taught and the primary responsibility was administration, the individual was considered an administrator. Department chairs in one public secondary school in this study have the equivalent of half-time teaching responsibilities with the remainder of their time being delegated to administrative duties. These individuals were labeled teachers for purposes of this study. In any event, the difference between teachers and administrators in this work proved not to be crucial, since each reported willingly and convincingly on their own actions and beliefs and those of their supervisors or subordinates.

Teachers were selected by a purposeful sampling from among those who were considered to be well-situated informants. For the purposes of this study, teachers who qualified for participation were those who had at least five years teaching experience and at least three years experience in their present school. Another criterion was to have a variety of subject areas represented. Each principal or head of school was asked to prepare a roster of full-time teachers from which the sample could be drawn. The interviews were conducted in the spring of the school year and pressures that naturally occur in all schools as they prepare for graduation and final exams precluded any scientific selection of interviewees. The reality of school life meant that principals or heads either asked for volunteers at faculty meetings, through department heads, or asked particular teachers if they would be willing to participate in the study. It is unlikely that a more scientific selection of interviewees would have resulted in any important differences in the outcome.

Data Collection

Good research practice obligates the researcher to triangulate, that is, to use multiple methods and data sources to enhance the validity of research findings. Mathison (1988) advised ". . . it is necessary to use multiple methods and sources of data in the execution of a study in order to withstand critique by colleagues" (p. 13). For this reason, multiple methods and sources of data collection were employed. Interview protocols were developed in such a manner that included, but of course was not limited to, questions from the High School and Beyond Survey (Moles, 1988) and Blase's (1991) study of power relationships between principals and teachers. It is the questions from the High School and Beyond (HSB) Administrator and Teacher Survey upon which Chubb and Moe (1991) based their index of teacher and administrator autonomy.

An interview protocol was designed (see Table 1)to explore these and related issues and utilized open-ended questions and probes. The purpose was to elicit reflective answers that go further than the type of surface response typically produced by a mailed survey.

Table 1

Administrator and Teacher Interview Protocols

Administrator Protocol

1. Can you tell me about an incident that happened to you or someone you know in which your work life was influenced or shaped by the .... (then A through L below). For example, staffing decisions, budget allocations, scheduling of classes, how you deal with discipline or behavioral problems, decisions about pursuing advanced degrees, how parent communications are handled?

   A. superintendent
   B. department chair

C. school board
D. state or federal programs or regulations
E. North Central or AIS (Association of Independent Schools)
F. legal or judicial judgments
G. parents
H. professional organization with which you identify; teacher's association
I. inservice training or your own continued education
J. students
K. colleagues
L. colleges

2. Describe the degree of control and discretion you are able to exercise over each of the following activities:

A. establishing curriculum;
B. determining instructional methods used in the classroom;
C. allocating funds;
D. hiring new, full-time teachers.

3. My research is directed at a current debate in education. It is claimed (by Chubb & Moe) that:

> A). Private school teachers have greater autonomy to innovate, adapt curriculum and teaching to meet the needs of their students, and that in doing so they are primarily influenced by the students and the parents and not by school bureaucracy.
>
> Whereas:
>
> B). Public school teachers are subjected to a variety of influences and pressures that restrict their autonomy in meeting students' needs; among these influences are:
>
> 1.) state and federal regulations;
> 2.) unions;
> 3.) court orders or the threat of litigation;
> 4.) organizational rules called "bureaucracy."

What do you think of this? (Probe: For example, suggest he or she compare Private school X and Public school Y in respect to the above question. Or ask the interviewee to imagine Public school Y and Private school X being the same size, then how would teacher autonomy differ between them, if it would at all.)

Teacher Protocol

1. Can you tell me about an incident that happened to you or someone you know in which your work life was influenced or shaped by the .... (then A through L)? For example, your selection of curriculum materials, what you teach or how you teach, how you group students, how you deal with discipline or behavioral problems, how your classes are scheduled, decisions about pursuing advanced degrees, how parent communications are handled?

A. superintendent
B. department chair
C. state or federal programs or regulations
D. school board
E. North Central or AIS (Association of Independent Schools)
F. legal or judicial judgments
G. parents
H. professional organization with which you identify; teachers association
I. inservice training or your own continued education
J. students
K. colleagues

L. colleges

2. Can you describe, out of your own experience or that of someone you know directly, a creative attempt made to improve the classroom, teaching methods, the curriculum, or student achievement that was thwarted or substantially altered by any of these (A-G in Question 1) sources of influence?

3. Can you describe for me a failed attempt by any of these sources to influence you that you resisted? (repeat A-G to remind participant of categories to consider) What are the ways that you have been able to work around those influences?

4. What does it mean to you when people talk about bureaucratic constraints on teachers?

Interviews were tape recorded and transcribed. Substantive field notes were maintained and reviewed at the end of each on-site visit and interview. All interviews were conducted by the author. Documents that were examined included Teacher's Handbook or Policy Guide, marketing brochures, school board and faculty meeting minutes, government regulations, and other printed matter deemed relevant to this study.

Data Analysis

Interviews were audio recorded and transcribed verbatim with dialog attributed to each speaker. Data derived from these extensive interviews, field notes, and documentation were reviewed for recurring themes utilizing the constant comparative method (Glaser, 1978; Bogdan & Biklen, 1982). Transcripts of the interviews were read repeatedly in a search for quotations that transcended the idiosyncrasies of individual circumstances and thus suggested a theme or idea about autonomy. It proved useful during the process to create charts as an aid to data reduction and analysis. Data display (Miles & Huberman, 1984; Wolcott, 1990) allows for the sorting and categorization of data in a way that seemingly discrete data may be linked in previously unrecognized ways. A grid was devised by which to organize and identify objects of influence on teachers and principals and the source of those influences. Developing themes were labeled and evidence was categorized accordingly. Then, quotations were extracted from transcripts and collected into files, with each file representing a distinct idea or theme. Quotations in files retained identifying codes that linked the quotation to its source interview. These "theme files" or categories were then read, edited and organized into a core set of ideas about teacher and principal autonomy. The core set of ideas was then reorganized by coalescing, splitting or eliminating themes until a satisfactory framework for reporting the findings was obtained.

The interpretation of categories became the basis for formulating a framework for conceptualizing the differences that may or may not exist between principals' and teachers' sense of autonomy and control in private and public school settings. The conceptual framework was used to describe the ways in which public and private school teachers and principals share a perception of autonomy, where they are different, and how they experience constraints on their autonomy.

In this study, the analysis of data and the reporting of interpretations are uniquely tied together. Below under Findings, each quotation that illustrates a concept of the interpretative framework is hyperlinked to the transcript of the interview in which it appears. By clicking on the icon beside each quotation, the reader can move to the quotation in the context of the full interview from which it was extracted. This feature of presentation of findings allows a check on the interpretation by the reader. In addition, the full text of all interviews is available to anyone who wishes to reanalyze the original data.

Return to the Table of Contents

The Schools

Three independent secondary, college preparatory schools and three public secondary schools were studied. Four of them are located in one large city, two in another. Independent high schools are defined as those being non- religious and, in this study, college preparatory.

St. John's College Preparatory School

St. John's is an independent, coeducational, college- preparatory school located in a quiet neighborhood in a city of nearly a half-million population in the southwestern United States. The grounds were originally a residence in a neighborhood more than a century old, although the school itself did not open until 1980. By a quirk of geography, the school is now located in an area marked by the influx of affluent families. The population of St. John's is more than 90% Anglo. The school has devised programs to entice greater student diversity; these efforts have met with little success. It has increased its Hispanic population to nearly 10% although the city in which it is located is as much as one-third Hispanic. Once recruited, many students tend to remain. Over half of last year's graduating seniors completed all four years of high school at St. John's. The withdrawal rate is low, only 1 to 2% annually.

As of the 1990-91 school year, there were 22 full-time teachers (11 women and 11 men) and 10 part-time teachers (five women and five men). The school reports that 23 of the 28 members of the faculty have masters degrees and three hold doctorates. The ratio of teachers to students is one for every nine. Full-time teachers are required to teach five classes. The teaching loads average between 63 and 65 students per teacher. The head of school held this position at St. John's for four years. He has over 25 years of experience in similar positions at three other independent secondary schools. He holds both a bachelors and masters degree from a prestigious eastern university. The position of head of school is described in St. John's self-study in part as one who "administers the school according to the policies set by the board of trustees" and has "complete authority for faculty, staff, and student selection, evaluation, and dismissal." Other administrators include an assistant head and dean of students. Both report directly to the head and, upon request, to the Board of Trustees. The head of school fulfills both the superintendent and principal functions of a public school system.

Tuition for the 1991-92 school year is listed as $6,650 for high school students. Students will incur additional costs for books and certain special events or trips. Some share of the funds, raised through annual giving, is allocated by the board to provide financial aid in the form of need-based scholarships. Student admissions decisions are made by an Admissions Committee. Qualifications for admission are based on admissions test scores (the Stanford Achievement Test is used), a transcript from previous schools attended, former teacher references, and a personal interview. No precise standards on these criteria are publicly stated.

Verde Valley Country Day School

Verde Valley Country Day School has the longest history of any independent school in the state. It began as a ranch style boarding school in the thirties, then closed its resident department and changed its mission to that of a four-year college preparatory high school. Verde Valley Country Day School is a non-sectarian, co-educational, college preparatory day school for students from grades 4 to 12. The school is located in a residential community which has seen the more affluent families migrate to newer areas of development within this growing city of nearly a half- million population.

Verde Valley Country Day School has a total student population of 188 for the 1991-92 school year; 105 of those are high school students. The student population is predominantly Anglo despite efforts to increase the ethnic mix in the school. The board and administration wish to attract a student and faculty population more representative of the city in which the school is located. All three of the independent schools in this study publish a lengthy list of elite colleges to which their graduates have been accepted.

Verde Valley Country Day School's full school faculty totals 24 full-time and 12 part-time teachers. Many of its high school faculty also have teaching responsibilities in the Middle Level School as is common in the other private schools in this study. Most teachers hold a bachelors degree; many hold one or more masters; a few hold PhD degrees. The administrative staff consists of the head of school, director of admissions, director of development, business manager, and assistant head. As is the case with each of the private schools in this study, the head functions as both superintendent and principal when compared to the public school system. Verde Valley Country Day School's Self Study states: "It should be noted that the key function of a head in an independent school is raising money and there is no question regarding this duty in either assessment [of the head in superintendent or principal roles]."

Tuition for all grades at Verde Valley Country Day School is $6,670 per year. Additional expenses include books, various field trips, a week-long program held each spring, and bus transportation. Financial assistance is allocated by a committee and based on need. Published information from the school states: "Each year a substantial portion of the school budget is allocated to financial assistance, making such assistance available to more than 35% of the student body, with grants in aid ranging from 90% to 7% of tuition for the year." Students are accepted for admissions to Verde Valley Country Day School on the basis of transcripts and standardized test scores from their previous schools, an aptitude test administered by Verde Valley Country Day School, recommendations from previous teachers, and an interview.

Crestwood Country Day School

Crestwood Country Day School is an independent, co-educational, college-preparatory day school for students from Pre-Kindergarten to grade 12. Of its total enrollment of 575 students, 169 are enrolled in secondary school. The school was founded over thirty years ago at its current location in an affluent residential suburban area imbedded in a city with a population of approximately two million. It enjoys a long and distinguished reputation in the area and is the only independent high school in or adjacent to that city. It competes only with public and parochial schools for its students and is generally viewed as an elite school by the community. Unlike the casual attire worn by the administrators at St. John's College Preparatory School and Verde Valley Country Day School, the male administrators are often seen wearing ties and jackets; the female administrators customarily wear suits. The philosophy of Crestwood Country Day School, as expressed by the assistant to the head, is one of a non-profit business accommodating the needs and wishes of its clients. The assistant to the head is also the director of admissions.

Crestwood Country Day School's student population is primarily Anglo although programs of financial aid exist to encourage a more diverse student body. Graduating classes average 40 students each year with 99% enrolling in four-year colleges immediately. Eighty percent of those students attend colleges out of state. Crestwood Country Day School is proud to include in its marketing materials a lengthy list of elite colleges to which graduates are accepted

Crestwood Country Day School employs 20 academic faculty in the Upper School. It is unknown how many of these teachers are full time. Masters degrees or higher are held by 80% of this faculty. Each administrator, with the exception of the head of school, is required to teach at least one course. The result is a teacher-student ratio of one to nine. Some of the administrators who are responsible for the Upper School serve the same function for the entire school. For example, the director of admissions is responsible for managing admissions for pre-kindergarten through 12th grade admissions. The top administrator is a woman with previous experience in the same position at eastern private schools who prefers the title "Head of School." She has been head of Crestwood Country Day School for seven years. There is a director of admissions who also has responsibilities as an assistant to the head of school.

Tuition for Crestwood Country Day School's Upper School students is $8,500. Students are also required to buy their books and may incur fees for special activities. The Board of Trustees has allocated a portion of the school's operating costs to provide an active need-based financial aid program.

In its marketing materials, Crestwood Country Day School clearly states that it "maintains a policy of selective admissions, recognizing that there is a range of students the School's programs serve best." Students are admitted based on an interview, testing and, presumably, other unstated criteria. Admissions decisions are made by the director of admissions and his assistant. The head chooses to participate when certain cases are considered.

Sunset High School

Sunset is a large, modern public high school housing approximately 2400 students. The state department lists the graduation rate at 83% with one of the ten lowest drop-out rates in the state. During the current school year, eleven National Merit finalists were identified. It is known in the community for having high academic standards and successful students. The community is primarily Anglo and affluent. The principal describes the student population as being 99% Anglo and possibly 1%, no more than 2%, other which would include African-American, Hispanic, and Asian students. While the district is largely middle to upper middle class, it contains small pockets of lower middle to low income families. Among the goals outlined for the school in its Teacher Policy Handbook is a commitment to "make a significant contribution to the needs of college- bound students . . . " and to maintain effective communication with students and parents "to best meet the educational needs of each student."

The superintendent has earned a national reputation for effective management and promoting academic excellence in the schools. She works well with the Board and the very active local Teachers Association. Policy is set by the Board with the guidance of the superintendent. The superintendent is then charged with the implementation of those policies. The implementation of these policies within each school is delegated to each principal. Teachers in the classroom are thus distanced from the Board by these layers of administration. The Board meets twice monthly to carry out their charge much of which is prescribed by the State Department of Education and applies to each of the public high schools in this study. Policies of the district are to be in harmony with state statutes. The state also determines the number of board members (five), term of office (four years), assumption of office, and fiscal year. Many of the board duties involve financial decisions, property management, and personnel discipline issues.

There are about 100 certified staff on the faculty of Sunset High School. They teach five classes a day and are limited to 160 contacts [students] per day. This means they average 32 students per class. Most teachers have been with the school since its beginnings, many have served in the district throughout their careers.

Portales High School

Much like Verde Valley Country Day School, Portales suffers from the migration of many of its more affluent families to the suburban areas of the city. It still retains a reputation for high academic standards, high graduation rate, and a significant number of National Merit Scholars. The community is facing a widening encroachment of lower middle to low income families, many of whom speak primarily Spanish. The student body reflects the community. It is about 90% Anglo, the remainder being primarily Hispanic and a few African-Americans. There are 980 students attending Portales in the 1991-92 school year.

The roles of the superintendent and board are identical for Portales, Sunset, and Montevideo High Schools. Their responsibilities of each are mandated by the state.

There are 43 faculty at Portales; about 55 if librarians, counselors, and part-time teachers are included. It is a mature staff, most of whom have worked within the district, if not the same school, for much of their teaching careers. Many hold advanced degrees as is prevalent in public schools where salary is tied to both years of experience and continuing education. Teachers are responsible for five classes and have between 25 and 30 students per class. The student-teacher ratio is 25.5 to 1. Portales has recently embarked on a move toward site-based management in which the principal encourages shared decision-making. Decisions, including the hiring of new faculty, are made by a group or team of those who have a stake in the outcome.

The current principal has two histories with the school, the past and the current. He had been principal from 1979 to 1987 including the period of the closure controversy. It was a time when the parents clearly made a choice for their community school. The principal left for a district level administrative position only to be recruited back as principal when his successor was removed under unclear circumstances. Most teachers report the removal was due to his inability to work with faculty in shared decision-making, others state a legal controversy around athletics. While the current principal appears ambivalent about his return to Portales, it is apparent that his quiet demeanor and stated trust in the competence of the faculty endears him to the teachers.

Montevideo High School

Although Montevideo was established as a high school seventeen years ago, it was housed within an existing district high school for its first year which required conducting a year of double sessions. The first graduating class, therefore, did not graduate from the present facility. After that first year, Montevideo has been housed at its own facility on its own campus. Its current principal opened the school and has remained with it until his retirement scheduled for the end of the current school year.

The school community is composed of middle to upper middle class families. The current student population of 2750 reflects the larger community, essentially Anglo and either LDS (Latter Day Saints or Mormon) or Catholic. Approximately 93% of the students are Anglo, the remainder include a small number of Asian, African-American, and Hispanic students. The graduation rate is over 90%. The current school year has produced eleven National Merit Scholars. The district spends approximately $3400 per student.

Early in the school's history, the assistant superintendent of the district began a study of the community by asking, "When your students graduate, what do you think they should have learned?" The school was, thus, established as an outcomes-based school from the beginning. Parent expectations have been reaffirmed three times since this first study utilizing a survey of parents and students. Through the survey, parents not only described their expectations, but they also ranked them in importance.

While the principal has complete control over his budget, he dislikes this role and prefers being an instructional leader, working with teachers and parents, and conducting long-range planning. He claimed almost unbridled autonomy in his work noting, "If you produce a good product, they're [district] going to leave you alone."

There are about 125 faculty who are described by the principal as "damn good faculty . . . intelligent, dedicated." They give much of their own time for tutoring students and they all incorporate mastery learning. They teach five classes and average 142 students within that teaching load. Only five faculty members were assigned to Montevideo, the remainder were essentially hired by the principal. Similar to each of the public schools in this study, the faculty is mature with many teachers having accumulated their teaching experience within the district, if not within the same school. They run their departments much in the same way one would expect to observe in a business.

The principal demonstrates confidence in those he has hired and allows teachers a tremendous amount of latitude. Teachers feel his support and express considerable freedom. In anticipation of his retirement, teachers report a general feeling of anxiety over the question of who will replace this principal.

Alike and Different

The six schools that participated in this study share some important characteristics, yet differ in a number of equally important ways. All serve primarily middle to upper-middle class clientele who are racially homogeneous. This characteristic is a result of two distinct factors: public school locations in relatively expensive residential communities and high tuition charged by private schools. All six schools are oriented toward college preparation. Each boasts of high graduation rates with large numbers of students accepted to both in-state and out-of-state colleges, including many elite colleges. Many students are academically motivated and earn academic recognition on both the state and national level. Parents are actively involved in the school, participating on the school board or on any number of committees. They are welcomed on each campus and their voices are heard. The parents are described by each principal or head as being generally well-educated and, consequently, understand how to get what they want through either the private or public school system.

The schools are remarkably different in size. The three public schools are all much larger than the three independent schools in both the size of the campus and student population. It is perhaps because of size that the public schools include a district level in their organization. Another difference is the amount spent to educate each child. Per pupil spending varies between public ($3400, $3700, $4000) and private ($6,650, $6,670, $8,500). While the public schools are forced to function and provide educational services to its students within its means, the private schools depend on fundraising to supplement tuition so as to cover the true cost of educating each of its students. Public schools are mandated to provide an education to all students, whereas private schools are selective of students and their parents.

Return to the Table of Contents

Findings: Teachers' and Administrators' Perceptions of Their Autonomy

This section presents a discussion of teacher and administrator beliefs about autonomy. In a subsequent section, teacher and administrator beliefs about constraints to autonomy are examined.

Teachers and administrators in both public and private schools reported, to a strikingly similar degree, a general feeling of autonomy. Teachers describe ways in which they experience autonomy: opportunities to participate in decision-making, support from the administration, and the ability to work around or ignore selected policies. School administrators also tell of having a sense of autonomy. Participants in this study described the effects of organizational size on their feelings of autonomy, how the administration acts to protect their autonomy, and the effects of teachers associations on autonomy. Their conversations brought to light the question of autonomy versus like-mindedness.

Teachers and Principals Experience Autonomy

I chart my own course through my pinball machine of life and I don't hit the bumpers unless I want to hit the bumpers.

Throughout the interviews numerous instances of expressions of autonomy can be found. Teachers in private and public school settings frequently expressed great difficulty, even frustration, in trying to rank the areas of control in their classroom work life. Participants in this study reflected Sedlak and others' (1986) contention that teachers today enjoy more freedom and autonomy than their predecessors (p.115). Certainly stipulations about professional codes of conduct of the 19th and early 20th century, sometimes viciously enforced by unbending administrators, are no longer the standard. The experiences reported by teachers in this study support the popular belief further pointed out by Sedlak and others (1986) that once they close their classroom doors, teachers are "able to exercise enormous discretion" (p.121). Public school teachers join private school teachers as they describe their sense of autonomy:

> I'll tell you what; we as teachers have a lot to say about all of these [items listed on questionnaire]. So, I would want to say that up front.
>
> It's kind of hard [to respond to the questionnaire] because I think . . . I could have number one, most control, on all of them; on every single one of them.
> [Click on the icon to the right to see this quotation in the context of the original interview.]

They are joined in their views by private school teachers:

> I'm very autonomous actually as far as my own classroom goes.
>
> In terms of my autonomy--you can see from my responses there--I feel a great sense of autonomy here.

Participation in the decision-making process.

Teachers in both public and private schools find expression of control through participation in curriculum or policy- setting committees. Many feel encouraged by the school administration to participate in decision-making; others feel they can participate by direct communication with the principal or head of school. Public school teachers described opportunities for participation in school decisions:

> In this district, anything that becomes policy has input from the teachers. . . . There are ongoing committees, and they are made up of a conglomerate of representatives. We have advisory boards in various areas. I think our school is one that utilizes teachers.
>
> What we have gone to is a system where the faculty itself has more of a hands-on approach to the administration of their particular program rather than going through chairs. Chairs still exist, but we are more of a local autonomy school now. We have a committee that meets and decides things with the principal.
>
> A public school principal reported:
>
> Everything that I do is a collection of information and input from teachers in this building, the department chair people in this building . . . They give me an awful lot of input. I'm constantly asking them for direction.

Private school administrators talked about how teachers are encouraged to create avenues of participation:

> The faculty have a big role in curriculum development . . . and usually it's a grass roots kind of thing . . . . curriculum change comes from faculty within the department . . . .
>
> There are many decisions that I'll just leave up to the faculty. The bottom line is that if you're going to have anything happen, you have to have the people who are responsible for enforcing it . . . part of the decision-making process.

A teacher in a private school typified the feelings of many in the small private school where faculty members feel and act like family members:

> If the headmaster does something which offends me, I go to the headmaster and we work it out.
>
> Yet this same teacher allowed,
>
> I would say the majority of the senior faculty are at a stage where they know even if he [head] doesn't like what I say I have a right to say it and he needs to listen to it . . . . Junior faculty might be a little too young to handle that.

Administrator support and encouragement.

Teachers in great numbers report they feel freedom in their work life because the principal or head or department chair has confidence in their expertise in content area and teaching skills. These are, in most instances, the very same principals and heads of school who hired those teachers in the first place. A school head tells how private school teachers gain autonomy:

> . . . [I] find very well qualified people with good imaginations to create good curriculum, and I support them and give them all the encouragement in the world to be able to do that.

Private school teachers described the atmosphere of autonomy due to administrative support:

> [We have] highly competent people. [The administration] lets them do their work and they either stay away by design or they are so busy they don't have too much time to get involved.

> I just really feel that she [department chair] has confidence in me and I have a pretty free reign.

Administrators have confidence in those they selected to be part of the school family. Yet, some teachers admit that this confidence may be tenuous. Support is evident as long as there are no parent complaints. Autonomy and administrative attention are felt by these teachers, but with the caveat added by McNeil (1986): "as long as the school runs smoothly". A private school teacher confirmed:

> The headmaster's role here--I look on it that he is very encouraging, that the office handles details like the scheduling and that kind of thing, but as far as how I run my classroom, it is pretty much up to me. I have a feeling that if there were a lot of parental complaints, I know I would hear about it. As far as structuring my curriculum, my teaching methods, even the way I handle discipline, I am pretty much free--as long as the head feels that I'm effective in what I do.

Private school teachers respond to the invitation of their administrators to utilize their perceived autonomy:

> Every creative thing that I've ever attempted has been encouraged at this school and people love my ideas and I've tried some pretty, you know, some things that I'm taking some risks doing.

> I get to design the whole course for the year of what I'm going to do in my classroom myself. I mean people know I've got a body of knowledge that I'm an expert at.

> A public school teacher added:

> He [principal] really relies on the department chairs . . .and as a department chair, I rely on what the teachers in my department want. It's a lot better way of communication and they feel like they have input; I feel like I've got input.

> Another public school teacher shared:

> All the principals that I have had have trusted me as a professional to handle my professional work the way I see fit. I have never had anyone tell me specifically what to do.

Ignore selected rules and regulations.

Despite what might be construed as constraints imposed by the larger bureaucracy of state departments of education on public schools or large district level administrations, public school teachers frequently maintain autonomy. They take control by ignoring, working around supposed constraints, or using what Sedlak and others (1986) refer to as "passive circumvention" (p.120). Using such methods, teachers are able to experience freedom in a bureaucracy that is unable to monitor actions or provide consequences for offenders. Kozol (1981) found that "imaginative teachers . . . have used their ingenuity and skill in order to arrive at a way out [of following mandates]" (p. 51). Indeed federal, state, and district regulations may operate in such a way as to limit the range of possibilities available to teachers; but, as Ball (1987) noted, "They certainly do not exercise absolute control within that range" (p. 247). A public school department chair stated:

> His [district administrator's] proposal was to decrease failure rates by changing the syllabus, by changing what we do. Of course, this is one we would love to mount the barricades for, and I side-stepped it at this school . . . by finding a creative way to enhance student performance in a real sense. . . . something called an Algebra Homework Initiative. . . . It reduced our failure rate by about 50%. It really side-stepped the issue of failure rate without diluting the curriculum to accomplish it.

Another public school teacher related:

> Individual teachers pretty well make up their mind as to which text they are going to use. The state has a list that they give out to districts. The department discusses the different kinds of textbook . . . . but individual teachers [make their own choices]. I teach from an entirely different textbook than my fellow teachers at [the other high schools in the district].

The same teacher went on to discuss the effects of a state mandated curriculum:

> The coursework that you are to teach and the other requirements that you have to have by law are really minimal. . . . You have the standard things you go by ... but for the most part, it is pretty much that you do your own thing.

Even those teachers in private schools not subject to the same government mandated policies as apply to public schools, also find themselves in the position of ignoring or working around school policy to preserve control of their work life:

> There are certainly plenty of rules and policies that I don't agree with, but very often I just ignore them. . . . in the faculty handbook, teachers are supposed to wear shoes, not sneakers. So I wear them [sneakers] and nobody says anything and that's that.

> One time they [school administrators] imposed an in-service program on us. We behaved so badly they have, since then, let us determine what goes into them. So I would say currently we have a great deal of control.

> I resist bitterly and strongly changing my teaching style . . . I resist and I do it either overtly by speaking out--expressing it; or, if that fails, one can very simply do it covertly in the classroom. Simply not do it.

This last response is exactly what Ball (1987) referred to as "omissive action;" simply not to do what one is instructed to do (p. 268). The teacher stated the obvious fact that behind the classroom door is where the greatest teacher autonomy exists, whether public or private.

Principals, heads of school, and other administrators also speak of feelings of autonomy. Although they may admit to sensing pressures from an administration or state regulators from above or parents, they perceive themselves as charting their own courses on behalf of the faculty and students to whom they acknowledge a responsibility. Both public and private school administrators who participated in this study were mature individuals. They have many years of educational administration experience behind them and understood how to work within their given system. They know how to make the system, public or private, work for them. In this sense, they were able to express a great deal of autonomy. A public school principal spoke about his feelings of autonomy:

> There are always parents in asking for this, asking for that, wanting this . . . . I work personally on a scheme of a frame of reference that does not let or works at not letting people impact me. It's my own personal--my wellness program of "I'm not a yo-yo and I'm not a pinball machine.

> Heads of school firmly stated these convictions:

> It is our responsibility to be service-oriented and to be responsive to our parents, but it is not our responsibility to place them in the position of calling the shots. . . . Our job is to please, our job is to serve, our job is not to allow parents to run the school.

> You can say I have a lot of authority and it would be a great deal on one hand. On the other hand, one could say the teachers have a great deal [of autonomy] determining what the curriculum is.

There seems to be a conflict of ideas here. If teachers are given a great deal of autonomy on issues of curriculum, hiring of faculty, and other policy issues, the autonomy of the principal is eroded. Powell (1990) questioned the compatibility of the an empowered principal who is to function as a leader and site-based organization which empowers teachers. He suggested that one must forfeit some degree of autonomy for others to become empowered. Yet, as Apple and Teitelbaum (1986) found, within Weick's model of a loosely-coupled organization different types of professionals can retain control and authority without changing or being changed by the decisions of other professionals. Teachers in public and private schools conduct their individual classrooms as they see fit without reducing the autonomy of the principal.

Organizational Size and Autonomy

In the public school we have a bigger organization so there may be more levels of bureaucracy because there are more people involved.

Both public and private school teachers and their corresponding administrators describe a work life with few constraints on their autonomy. Common sense, however, dictates a focus on some obvious differences between the public and private institutions which create different reasons for a feeling of freedom. An obvious difference between the public and private secondary schools in this study is their size. Montevideo, Sunset, and Portales High Schools have student populations of 2750, 2400, and 980, respectively. The independent schools have populations of 275 (St. John's College Preparatory, grades 7 - 12), 104 (Verde Valley Country Day School, grades 9 - 12), and 169 (Crestwood Country Day School, grades 9 -12). Questions of size, who gets hired, the role of the principal and head of school will be discussed in terms of how public and private school teachers acquire autonomy.

Some of the autonomy in curriculum matters reported by private school teachers derives not from the organizational structure but from the fact that private schools are small requiring fewer demands for cooperation and coordination among teachers teaching the same subjects. In this section the necessity of a standardized curriculum to maintain continuity in large districts, layers of authority required of large organizations, and response time will be discussed. These issues are matters of the size of an organization that distinguishes between a public and private high school.

Curriculum decisions in large schools require discussion among the department faculty. The math department at Sunset High School, for example, has a faculty of sixteen. Faculty representatives in each content area pursue curriculum discussions with their counterparts in the other high schools within the district as well as coordination with the middle schools which send students to the high schools. There is a close articulation of curriculum to preserve continuity in both the scope of a subject and its sequence. Teachers influence curriculum through participation on curriculum and textbook selection committees.

Contrary to Lortie's (1975) description of the isolation and separation of teachers into the eggcrate conception of teaching, teachers in modern high schools have centrally located conference and work areas. Each of the public schools in this study had such a meeting area available for each subject area department. It is in these areas that teachers held department meetings, met with students, conferred with parents, collaborated on instructional and student needs, and prepared for instruction. A number of interviews were conducted in rooms of this type. The small size of the private school precluded a convenient area for teacher collaboration. A combination workroom and faculty lounge was where teachers could meet and confer unless a classroom were available.

A new teacher to the public school is expected to build his or her course around a given district curriculum to maintain continuity among the schools of a large district. While the public school expects teachers to follow the district curriculum guides, they are just that-- guides. An established curriculum does not mean there is no room for innovation. The presence of a curriculum does not deny creativity. An assistant principal of a public school stated:

> I think it came out when we had district-wide curriculum meetings, when the high schools were talking to middle schools and other high schools and we sat in rooms made up of representatives of the various schools. We talked about their relationship in the curriculum. I think there was discussion about the rigidity and that you shouldn't impose this upon teachers, but let teachers be more creative. I think that discussion was there and I think the realization was there that you also are tied in to some curriculum guide.

Teachers in public schools talked about how a district curriculum does not constrain autonomy:

> On the district level we have curriculum that we must follow. . . . there is no specific pressure or anything like that, but in a district the size of [ours] you have to have some coordination and articulation. . . . we have committees that work out curriculum problems, et cetera and select textbooks . . . we are expected to abide by those guidelines. But I don't consider that to be something that has come from on high. That is something that is logical. You would want all the schools in one district to basically follow the same core curriculum, but the core curriculum is only meant to be about 60% of the curriculum. Forty per cent of the curriculum we can decide on.

In a private school, new teachers will generally define the curriculum predicated on their own content knowledge and interest. Because of smaller faculty numbers, there may be two or three other teachers with whom to coordinate curriculum; yet each teacher specializes in a particular facet of that content area. While each of the three independent schools in this study have either a middle school or middle and elementary school as part of its organization, students come from a variety of other schools. Consequently, coordination is a matter of interest only within the upper school. Any coordination of curriculum is accomplished within the institution, as described by this private school teacher:

> I think we're all on the same track, which you might attribute to the fact that it is a small school. It is a college prep school. They're [students] all basically going through the same thing, and that certainly could be a strong positive as opposed to a larger school, particularly a large public school where you're serving many, many different peoples and one of those might be the college prep oriented students.

It was during a discussion of size of the institution and teacher autonomy that the head of a private school stated:

> I stress that not only can they have the pleasure of a great deal of autonomy here, they have the responsibility of it. No one will hand them a course outline and for some candidates that's very uncomfortable . . . . They'll even say, "You mean no one will tell me what book to use and what materials to use?"

Layers of bureaucracy appear to be necessary for the functioning of large districts and large high schools. A principal of a public school plainly states, "In the public school we have a bigger organization so there may be more levels of bureaucracy because there are more people involved". Despite the large size of the public schools, autonomy need not be compromised as confirmed by many of the public school teachers and principals in this study. It is because of size that the department chair functions in a role similar to the principal in terms of leadership and support. The department chair involves teachers in decision-making and communicates their position to the principal. The chair can also be another buffer to protect teachers from external pressures as will be discussed in the following section. It is size that requires teachers to work together, as these public school teachers reported:

> The principal has picked department heads that are facilitators, that can help that department be cohesive and bring out the best in the people there. . . . he [chair] has an interest in everything and can build a rapport and make this a cohesive, dynamic group. No one is ever stuck with all the dribble courses. You know, we always laugh, "Into each life some freshmen must fall."

> I don't have a lot of department meetings because I'm always seeing them . . . . I teach three classes and because it's such a large department, I can get out the rest of the day and be with them. I'll be in the classroom and I do most of the observations. I'm in the classroom even if I'm not observing, and that's when you really see what's going on anyway.

It is generally acknowledged that size slows down the response time of problem solving or making changes in policy or curriculum. In a public school there often is a hierarchy to be accommodated: one or two levels of administration, perhaps the school board, committees, and others from whom response is necessary. A comparison between public and private school life was made by a public school principal who had former experience as a head of school:

> I get frustrated here sometimes in that between the conceptualization of an idea and implementation it takes time; but the danger of the [private school setting] is that you are relying entirely on the head to make all those decisions. . . . It is not always so good . . . I'm not always right and sometimes I make mistakes. I think sometimes it's better if an idea is looked at carefully, if it's bounced off other people . . . but I don't feel in most cases that our classroom teacher performance is held back by that.

> This same principal of a large public school states:

> Things that hold back the classroom teacher performance probably deal with other factors to me. One deals with class size. . . . When I see the teacher too busy to go back and spend a few minutes with one, two or three kids, that's a problem.

Principals Protecting Teacher Autonomy

I guess that's the one thing about my department head, my principal, my superintendent; they don't crumble when there's a cranky parent.

Although size of the institution plays a primary role in the perception of quality, the role it plays in the autonomy felt by public and private school teachers and administrators is more complex. The roles of the principal or head, superintendent, school board, and department chair; teachers' association; and the determination of who gets hired all contribute to the sources of autonomy that can be found in schools.

Contrary to the beliefs of some, administrators in both private and public schools often act more as buffers protecting teachers from pressures from outside groups than they act as sources of pressure themselves (Blase, 1991, p. 736). The image of the non-supportive administrator who saddles teachers with trivial tasks and burdensome paperwork (Boyer, 1983, p. 142) was not found among participants in this study. Nor was there evidence of the type of principal that talks at and delivers commands to teachers or staff meetings that concentrate on administrative details ignoring matters of educational policy as described by Boyer (1983, p. 224).

In the private school, the role of the heads is such that they act as both superintendent and principal. They determine the philosophy of the school and train the board as to their policy making and fiscal responsibilities. The head or superintendent, once hired by the board, is charged with seeing that the school board or board of trustees separates policy making function from that of the principal or head who sees to the daily management of the school. When heads or superintendents do their jobs well, the teachers feel no constraints from the school board or board of trustees. Teachers in both public and private schools generally agreed that the board "stay[s] out of the daily running of the school," as stated by a teacher in a private school. Another private school teacher opined:

> There's a layer between me and them [board of trustees], and that layer is [head] and [assistant head]. . . . You know, I might be doing some things which are driven by board decisions and I just don't know it.

> The head of a private school added:

> We don't have an education committee on the board. I view an education committee on the board as potentially dangerous because, in fact, there was one when I came and I let it die . . . that is an area where they can easily lose sight of their responsibility . . . when you have a formalized structure it can get dangerous, as

> opposed to an informal structure where just some parents are saying that would be great if we has this or that . . . once you formalize it, it can become a problem.

A public school teacher acknowledged how the principal worked on behalf of the teachers:

> Our principal was spearheading, and he did get permission of the board to do it, even though it meant working the system a little bit. It is a pilot program, but it's not being called that because, if it were called that, he would not be able to do it in the middle of the year.

Principals, heads of school, and department chairs are generally seen by both public and private school teachers as being supportive and protecting them from external pressures. Knowing these buffers exist allows teachers greater flexibility and freedom in their work life. Public school teachers commented on the support and protection their administrators provide:

> I guess that's the one thing about my department head, my principal, my superintendent; they don't crumble when there's a cranky parent. All the lines of communication are followed in a correct way, and I'm helped along the way. They don't give in to that parent, parental pressure when it's just a cranky person out there not getting their way. They're very articulate about it. They're very professional, but the buck does stop here with the department head, with the principal and with the superintendent.

A public school principal related this story:

> I recently went through hell, two weeks ago, with a mother and a father over a boy who didn't graduate and the parents were insistent that I graduate him. [They went] all the way to the superintendent level, bringing the assistant superintendent out here because we were not being fair with that kid. The teacher was being very fair with that kid, very fair, and I supported the teacher and the kid did not graduate . . . . They wanted the teacher to go back and change a grade and I'm not going to make a teacher do that.

Teachers in independent schools described similar feelings of support:

> I think [the head] screens and keeps us away from parents who would stop some program. He very much wants the teachers to have the feeling of freedom to teach whatever they want to.

> Basically what he said and I've heard him say publicly is that we aren't going to change our curriculum to suit an unhappy parent. We're willing to look at our curriculum and see if it's what we ought to be doing, but we're not going to be in the position of, you know, changing because a parent is unhappy about something. So we have a lot of support for that.

A private school administrator responded to parent pressure to fire a teacher:

> You positively get a lynch mob going in a situation because in the second week we had people calling us to fire this woman . . . we toned them all down and even some of the other parents would say to the rabid parents, "Isn't it fair to give her a little time to get adjusted?"

Teachers Associations Affect Autonomy

Our teachers association is very active and it affects my work life every day.

There is an acknowledged criticism of teachers associations in the realm of public opinion and among critics of public school systems. This study was conducted in a right-to-work state in which teacher unions are virtually non-existent, but teacher associations are predominant. These associations are seen as variously strong or weak depending on locale. Only one of the three public schools is in a district having a very strong teacher association. Most, if not all, of its teachers are members of the association and quite a few are active in its leadership. The other two schools are in districts that negotiate teachers' contracts with the association, although the faculty are much less active. Teachers in all three public schools, however, reflect on the efforts of the teachers' association to preserve their autonomy. If educational researchers (such as Chubb and Moe) promote autonomy as the key to freeing teacher creativity and innovation, they should applaud the efforts of the teachers association which acts to preserve the due process upon which teachers have come to depend for a sense of freedom in their work life. It is the teachers association that can require a district to seek advice from teachers, to protect teachers from pressures to change grades, and to provide good working conditions.

While the association does protect specified areas of teacher autonomy, it also institutes a management system based on the model of industrial unionization leaving many teachers feeling more powerless than before (Russo, 1990, p. 193). Despite the price they may pay, Firestone and Bader (1991) credit the teachers association for the extent to which teachers participate in program design within a school system (p.84). Both public school teachers and principals, who at times may feel constrained by the presence of the local teacher association, express positive reactions toward the association. One public school teacher explained:

> He [former superintendent] was dictatorial. It's this way because he would sit back and smoke his pipe and he would smirk at you, and his aim was divide and conquer . . . . I think that is when our association became the dynamic force it is because he was so bad and that was when the parents realized that there was a dynamic force out here called teachers, and their [teachers] main goal was good education, not paychecks. It was like we are your comrades, not your enemy.

A public school teacher who formerly worked in a union state on the east coast speculated about why unions or teacher associations are important in protecting teacher autonomy:

> . . . . and there we actually had more autonomy and I feel that way because it was unionized. . . . The only reason that I believe unions have ever appeared is because they had employers who are less than honorable and kind of impose their will . . . they were autocratic and we wouldn't have a need for an association or union if you didn't have individuals such as that.

A principal in the public schools said:

> The administration seeks their [teacher association] opinion. We let them know when decisions are being made that we think are going to have a significant impact on the faculty. . . . We include them a lot, we treat them as equals, we value their judgment and input, and I think there's a good working relationship.

The teachers association can interpose itself between the teachers and the principal and protect teachers from unfair or unjust decisions (Grant, 1988). One public school principal stated:

> I have a reputation for dismissing teachers, that I'm Atilla the Hun, if you will, about evaluation and I am. People will tell you, "You can't do that with a professional organization." My organization works beautifully with me because I dot the i's and cross the t's and I treat the person humanely as I'm doing it. Therefore, they never have grounds to come in and say you didn't follow procedure or you treated these people like dirt. As such, I usually end up with very strong support from them.

Autonomy or Like-Mindedness?

I find very well qualified people with good imaginations to create good curriculum, and I support them and give them all the encouragement in the world to be able to do that.

Much of the autonomy felt by teachers in each school derived from the fact that they were in agreement with their administrator. Principals and heads hire teachers who agree with their philosophy. It is only on occasion, with declining school enrollments and concurrent reduction in teaching force, that a public school principal is forced to accept a possibly unwanted teacher on transfer from another high school within the district. Otherwise, they feel great control in selecting new teachers.

In both the public and private schools, the principal or head screens the potential teacher candidates before seeking advice from the faculty. At times the teachers in the private schools in this study had to fight to participate in the hiring of new faculty. Perhaps heads are less willing to share the task because their jobs rest on the selection of teachers who must be perceived by parents as effective to maintain the school's very existence. A head of school relates:

> . . . [I] find very well qualified people with good imaginations to create good curriculum, and I support them and give them all the encouragement in the world to be able to do that.

Retaining control over the hiring of new faculty for both the public school principal and private school head ensures a faculty with a philosophy shared by the administrators. Teachers expressed their consternation over being left out of some aspects of the hiring process. These two private school teachers described their role in hiring colleagues thusly:

> We're in the process of hiring a new teacher. It's been quite a frustrating experience. . . . I am not allowed to see recommendations, but I am the art department chairman. I have interviewed several candidates. I have looked over 85 resumes for this job, and I've yet to see one letter of recommendation. I don't know, I've never been told . . . . Apparently now the only person who sees them in this school is the headmaster, and one other person-- and I find that to be a little degrading.

> They began the process of hiring a new drama teacher, reading through resumes and inviting some [candidates] without ever letting me know as head of the fine arts department that they were considering this person. And you don't do that. You don't do it. Well, I went in and jumped up and down and raised holy hell and the response there was copious apologies giving me the resumes to look at, asking for my opinion.

Knowing that faculty view education through a similar set of beliefs, principals and heads can comfortably allocate greater autonomy. They can give support and show trust in teachers with the knowledge that teachers are "like-minded." Also, as long as a school, public or private, is perceived by the community and parent body as successful the principal or head is less likely to interfere with teacher freedom. Perhaps the issue of autonomy is derived from the principal or head and faculty acting in ways that have the approval of the parents. Knowing what parents want and sharing those expectations translates into autonomy for teachers. The support of parents adds, as well, to the principal or head's autonomy. Private school teachers talked about fitting in at their schools:

> In my last school where I worked my department chair caused me to be fired. . . . If it matters, I'm much better [off] here than I was there. I mean I was a square peg in a round hole there and here, it's a much better fit.
>
> I'm pretty much free as long as the head feels I'm effective at what I do.
>
> A public school teacher reported:
>
> We were rolling along at this school. This school was a great school, and it was because of the teachers. We were heading in the right direction and so on, but the difference that I see is that he [principal] has come in and given us some direction, come in with some new ideas. The ideas we had before he has improved upon, given us freedom to do these things.
>
> A head of school described his hiring practice:
>
> I'm the one that will usually go through all the applications, bring it down to about ten, call them in, interview these different people, then I make the final three selections. Then I'll bring in at that point the department head or a couple of other teachers . . . . You know, it's generally my decision almost alone.

Public school principals reported having considerable freedom in selecting teachers. If a reduction in force is in effect in a school district due to declining enrollment, principals are required to accept transferring teachers. One principal explained that of his current faculty of 125, only about five were not of his choice. Another principal explained how the school organization is becoming increasingly more site-based. Department teams screen, interview, and hire new teachers for the department. The principal may be part of the team. Sharing the hiring process removes some autonomy from the principal, yet he has trust in the faculty to make good choices. Perhaps if there was a lack of trust, the process would be different.

Teacher Autonomy in Pubic and Private Schools Compared

> "I think we have fully as much freedom in public school as they have in the private school."

The teachers who participated in this study view themselves as active participants in making many of the decisions that affect their work life. They describe many opportunities to participate in and influence policy decisions. They talk about having control over what happens in the classroom even if they retain control by ignoring or working around bureaucratic constraints. One public school teacher described what many of his colleagues also believed: "Bureaucracies within our district are . . . as far as influencing what happens to me as a teacher, almost nonexistent".

Private school teachers also report having a great deal of freedom in the same areas, but attribute it to a freedom from state and federal constraints: "Being an independent school, we aren't bound by the required [state] curriculum . . . . I don't feel shaped by the federal government. . . . I feel very fortunate that I sense control in an inordinate amount of things here". Principals also talk about taking control and responsibility for their work lives. A public school principal reflects the views of his colleagues: "I think we have fully as much freedom in public school as they have in the private school". Heads of private school view their position as one permitting immense freedom and having the ability to define the roles of others who work within the institution: "It's always up to the headmaster to help educate people when they are overstepping their bounds." Another head reflects: "We have the autonomy to change a program entirely if we want to . . . ."

Teacher and Principal Autonomy, As They Tell It

Teachers and administrators in public and private high schools in this study feel that they experience a great deal of freedom in their work life. Equally evident is the fact that none can claim unrestrained autonomy.

Whether public or private, teachers' explanations for feelings of autonomy are similar. Participation in decision-making gives them a sense of influencing school policy. When they are encouraged and supported by the administration, teachers feel free to take risks in teaching and they adopt creative and innovative strategies (Blase, 1988; McNeil, 1986). Often when externally imposed rules, regulations, or mandates infringe on this freedom, experienced teachers and administrators ignore them or work around these obstacles.

Teachers and principals in public and private high schools also described three features of school organization that enhance and protect autonomy: (1) the size of the organization, (2) administrators acting as buffers, and (3) the teachers association. First, teachers and principals in large public schools find that factors related to the size of the organization help to protect and maintain autonomy. It is acknowledged that there is a vast difference in size of organization between public and private institutions. Contrary to the popular belief that layers of bureaucracy act as obstacles to autonomy, the organizational structure of large schools enhances autonomy by clarifying roles so that public school teachers are faced with less ambiguity. Within the role and within the classroom, teachers described a sense of freedom. Public school teachers are expected to work within the curriculum guidelines of the district and state, but are given broad latitude within which to innovate and be creative. Size also requires some standardization to accommodate articulation of curriculum content from middle schools to high school and between high schools of the same district. Private school teachers, on the other hand, may enjoy even greater freedom in that they often write their own curriculum. Although two or three private school teachers of the same subject may share ideas, there is little need for cooperation since it is unlikely any two of them teach the same course to the same grade level student. If one could imagine a private secondary school of two to three thousand students, it would likely function in much the same way as an upper middle-class public school with regard to administrator and teacher autonomy.

Second, autonomy is protected and maintained as principals and heads of school act as buffers to protect teachers from external influences. In public schools, the assistant principal and department chairs form additional layers that protect teacher freedom. Even the public school board can act to support teacher autonomy in the classroom. Heads of school do the same. All of these groups expressed, orally or in writing, a philosophy of management that shields teachers from external pressures. The board of trustees of a private school does not share this perspective since they are kept away from the daily business of running the school. Rather, the function of a private school board is to establish or approve policy and to raise funds, both roles lying far from the classroom door.

Third, public school teachers are given some guarantee of protection of working conditions by the teachers association. They cannot be subject to unjust firing. The association protects teachers in ways that leave them fearless in the face of some external pressures. For example, public school teachers cannot be pressured by parents or administrators to change a student's grade, the number of student contacts (number of students per class) is limited, and teaching responsibilities are often specifically delineated. The teachers association also negotiated mandatory teacher participation in decision making through committee work. Private school teachers have no similar protections though they are subject to few of the public teachers' concerns because of the size of the organization and the heterogeneity of the student body.

Findings: Constraints on Teacher and Principal Autonomy

Any attempt to clarify and elaborate the concept of autonomy would not be complete without an investigation of those pressures that act to constrain autonomy. Despite the strong sense of autonomy reported by those interviewed, they also acknowledged areas that compromise their autonomy: pressures exerted by parents of college-bound students, a context of laws that apply to both public and private institutions, financial constraints, and maintenance of an atmosphere which is responsive to parents. Public and private school teachers and administrators are often subject to similar, if not identical, constraints.

College Admissions Pressures

When you sign on for an AP, you're largely signing on to mandated curriculum.

Teacher autonomy in both public and private secondary schools is sharply compromised by the demands of parents wishing that their children gain admission to prestigious colleges. It is not uncommon in private schools to hear of parents and alumni wholly preoccupied with admittance rates to colleges (Lightfoot, 1983, p. 295). In its marketing materials, each private school in this study included a lengthy list of prestigious colleges to which their graduates have gained admission. The principal of each public school boasted a high graduation rate with many graduates being accepted at the best colleges. Each also expressed pride at offering a wide range of advanced placement (AP) courses and producing a number of National Merit Scholars. Parents in both the public school and private school communities are acknowledged by the faculty of each school to be highly educated, professional, and generally to be upper middle to upper socioeconomic in social class. It can be presumed that one of the reasons parents place a child in a college preparatory independent school or locating the family in a particular school district where the school has an reputation for academic excellence is the strong desire for the child to be accepted by a prestigious college. These parents are often actively involved in school activities or participate on committees. The demands of these parents are made known to administrators and teachers through direct contact or participation on school committees. Administrators may be more intrusive in this arena because the stakes are highest where parents are outspoken. Private school teachers described parent pressures:

> Occasionally you see parent pressures. Sometimes we have parents that are pretty pushy with their kids . . . we're dealing with some parents who are, you know, where both the parents are professional people and very busy and they essentially think that once they pay their tuition that you're going to take over dealing entirely with the student's education.

> Parents wanted that course [AP calculus] . . . . if there are enough [parent] voices behind there, it would have an effect [in making these curriculum changes].

> In college counseling, parents play a tremendous role, and they can put incredible pressure on me as a college counselor. "Johnny has got to get into college. I want you to do everything you can to get him in that school." And often people like that, and it doesn't mean just Harvard it can be Westminster College, will try to wield power over you. Again, it's [not] that you have to do this work, but, "I'm telling you how to do it," undermining in a sense maybe your professionalism, your training, your experience and expertise.
>
> This year AP class had to be geared to college expectations. I really had to adhere to what would be tested. In some ways, [I] lost some of my freedom in that class because I had to focus on college expectations.
>
> Private school heads and administrators similarly described parent pressures:
>
> Parents who send their children to private schools occasionally behave as if they owned the faculty, as if their amount of tuition were paying the faculty, each faculty member's entire salary.
>
> . . . if it [what a teacher is doing] also achieves all of our other goals for college preparation, things that we are trying to be sure we are doing for kids, we're able to allow more autonomy and we're able to try to work with parents in terms of informing them in a more cohesive way.

Pressure exerted by parents of college-bound students are felt and reported by public school teachers as well. Textbooks and curriculum choices are seen as examples of teacher responses to these pressures:

> I've been department chair now, it's been about seven or eight years. . . [principals] override specific decisions about placement into honors courses. . . . placement is not supposed to be determined by parents or principals, it's supposed to be based on certain criteria . . . . I should say at least once a year, principals override those decisions because of parental pressure.
>
> Our particular community here around [school] is very achievement-oriented most of the time, so there's a lot of pressure for kids to get good grades, and getting a B for a lot of students is a disaster. . . . I think there's pressure there to offer more AP courses because more and more parents are allowing their students to take advanced placement and try to get college credit before they get out of high school.
>
> Most of our kids talk college. We do have an academic program that is very heavy in that regard . . . more advanced placement classes being taught . . . a number of A level classes that would be appropriate for a kid going to a four-year or to a highly selective school; and we put a lot of emphasis on that, because the public is asking us to.

The public school principals in this study are very supportive of advanced placement courses and programs geared to the academically talented or college-oriented student. One principal boasted:

> We have the largest advanced placement program in the state . . . .When we began to excel in advanced placement and did a lot of publicity, [the superintendent]

> mandated that all the high schools in [the district] would have advanced placement programs. . . . . We're about the top three percent in the United States in advanced placement participation and success.. . . . we've had a remarkable run. I've had great influence that way.

A private school administrator added,

> ". . . when you sign on for an AP, you're largely signing on to a mandated curriculum."

College requirements and the College Board which produce the advanced placement exams influence public and private high schools to an equal degree. Parents of college-bound students in both public and private schools expect to have such courses available to their children. Teachers of core subjects, therefore, tend to look to these requirements when selecting textbooks and planning curriculum. All six high schools in this study contain college bound student populations. Preparing those students for college is a high expectation of parents and, consequently, a priority for the schools.

Since there is so much emphasis and concern placed on advanced classes in the core curriculum areas, it is interesting to look at how the teachers of non-college preparatory courses view their work life. Teachers in both public and private schools experience greater autonomy when their subject is not a college preparatory course.

> I may have more freedom than teachers in some academic areas . . . there is no set of standards and curriculum in the arts that high school students are expected to have by the time they finish high school. Therefore, I don't have anyone breathing down my neck to say, "You aren't doing this and this standardized test requires that you do that." So the subject area allows for considerably more freedom.

Threat of Litigation

We all feel the influence of lawsuits and insurance demands.

Autonomy of both private and public teachers is limited to an equal degree by a system of laws. Laws that have to do with civil rights, health, and safety are binding on the private institution as well as the public. These laws and the possibility of legal action compromise autonomy. Teachers have forgone some of their freedom knowing that lawsuits have only multiplied in recent years (Grant, 1988, p. 141). Heads of school explained how they are subject to the same constraints placed on their public counterparts:

> Any time that regulations come down through the federal government, it's pervasive in terms of health reasons, you know, it's pervasive throughout our society. We obviously have to adhere to those things. . . . We have to adhere to, of course, general health standards that exist in [the county] and the state. We test our water on a regular basis . . . we adhere to fire regulations; we have our fire drills once a month.

> We are subject to virtually all federal laws regarding discrimination. We publish a disclaimer in all of our publications stating that [the school] does not discriminate on the basis or race, creed, color, et cetera. A violation of that would and should mean that we, as an institution, should be closed or lose our non-profit status.

The teachers, principals, and heads of school were very aware of the threats of a litigious society and make conscious efforts to avoid such difficulties. Fear of lawsuits constrains public and private school decisions alike. Private school teachers reported how the fear of lawsuits has altered their work life:

> We all feel the influence of law suits and insurance demands than we used to . . . . It's that level of influence. I don't feel quite as free to do some things just because people sue each other these days.

One head of school described how he felt somewhat more secure in a small, private school setting than he would anticipate in a large, public school when it came to thoughts of being sued:

> Now it's not that we can't get sued as well [as public schools], but at a smaller institution you're more family-oriented. Things are based on more of a civil way of handling things, and you try to figure out how you're going to manage the problem other than just automatically jumping to think you're going to get sued.

External forces mandate and regulate schools and teachers so as to "provide adequate instruction to all their students, to equalize access to knowledge" (Sedlak et al., 1986, p.118). Ball (1987) reported, "The more diverse the school community, the more difficult it will be for any school to respond to all expectations" (p. 251). Even in public schools with little diversity, these constraints are experienced. The fear of litigation was felt by public school teachers and principals to the same degree as their private school counterparts:

> At the beginning of the year, we had a parent who came to us with an order from her attorney that they were going to proceed with bringing action against the district if, in fact, we did not change a grade that her son received because he was diagnosed late in the year as having attention deficit disorder and she felt that not every teacher did make adjustments in the teaching procedures to reach that child . . . . we met with the teachers a number of times and finally the teachers, out of a sense of inadequacy and frustration, felt that they did not want to go through a legal situation, so they changed the grades in some cases.

Financial Pressures

. . . we are becoming more like the private school, where the willingness to fund the institution determines its success

Yet another constraint placed upon both public and private schools is finances. Private and public schools are plagued to an equal degree by the shrinking value of the dollar and an unstable economy. The tax base upon which school funding rests is dwindling while the number of families who can afford a private education is stable, at best; certainly the numbers are not rising. The private school is also necessarily dependent on its fund-raising abilities; tuition alone does not cover the cost of educating each student. A head of school described the private school's quest for financing:

> Our [private school] burden is raising money. The tuition pays for maybe 80 percent of what we do and the other 20 percent we have to raise one way or the other; through fees or through fundraising or whatever . . . . they [public schools] don't have that same burden, although they have to go through elections and bonds and trying to get the public vote.

A public school principal dispelled the myth that public and private efforts to acquire funds are so different:

> We are attempting to work more closely with the community, with business and industry partnerships, things like that which is more like fundraising. It's more like what's being done in the private schools . . . . In that way we are becoming more like the private school, where the willingness to fund the institution determines its success.

Decisions on class size, the ability to offer additional classes and to purchase books and equipment are all dependent on the financial support available to each school. Some of these decisions are made by school boards and Board of Trustees, others are made by the principals, superintendents, or heads of school as they prepare their budget requests. The results affect the autonomy and work life of the private school teacher and the public school teachers in vastly similar ways. Frustrations in the private institutions were heard:

> Many constraints that we have are bottom-line dollar kind of restraints. . . . That doesn't mean we sell out to the dollar; it does mean, sometimes, that we have to give in or buy in where we would prefer not to. . . . because of the monetary factors alone, because of fewer people doing more different jobs, some of the autonomy is not quite as great as one would like.

And from the public schools came teacher and principal comments:

> The school board, two years ago, did away with a cap that we had on English class enrollment. We wanted to limit it to 125 students a day or 25 in a class and we'd had that cap for eight or ten years and because of budget constraints, they did away with that two years ago and now our classes are 30, in the 30s, up to 30, over 30. That's had a great deal of effect on us.

> Well, constraints, in terms of the amount of staff that we have, money becomes the bottom-line issue. If we could have five more teachers, we could have more and smaller classes.

Parental Expectations and Demands

I mean, one call [from a parent] in a district as large as this means a lot and that's just the attitude of this district.

Proponents of school choice often describe private schools as small businesses that must be responsive to clients, assumed to be the parents, in order to survive. Indeed, teachers and administrators in the private schools who participated in this study affirmed the expected and incorporated the language of business:

> You know, private schools are small businesses essentially, and you have to do business. The customers are the parents, but give them what they want, not as far as grades; don't give them the grades they want for their kids, but what the heck, if they want more feedback, they've got it.

> In a private school . . . you've got people paying $6500 a year to send their kids to school. You tend to appease parents a lot more than would ever happen in a public

> school. . . . a situation where perhaps I would have come down pretty hard on the situation . . . and if caught in that position, well, what do you do? You have to consider where your bread and butter is coming from.
>
> . . . you've got to be smart about it; you have to know how to market your school properly. You have to keep your customers satisfied . . . and you have to have good communications. So those are things one has to consider and therefore, parents are a very important aspect of the school.

Contrary to the myth that has been perpetuated by some, the existence of a bureaucracy does not necessarily imply insensitivity to the desires of parents. Indeed, there appears to be no lack of sensitivity to parents among public school teachers who report frequent and important contacts with parents, no more or less than occurs in private schools. Both teachers and administrators understand the expectations of the parents and make considerable effort to be responsive to those expectations. Some public school teachers acknowledged that limits to their autonomy are frequently set by community standards. If teachers are of like mind with the parents in the school community, they have a greater sense of freedom. If they do not, they feel constrained:

> I've changed the way I react to a negative parent. I think I tended to put them on the defense too much, and I'm like, well, "What is it that you want from me at this point? What is it that I can do to make your child be the best they can be?" . . . . I've learned that from [department chair]. . . . he makes them a member of the team rather than a member of the enemy.

The assistant principal of a public school explained the kind of response to parents mandated by the district:

> Parents' calls mean a great deal. We have a procedure here that if they're not satisfied with my answer, they can go to [the principal] who is very, very responsive and receptive to parents and if they're not satisfied there, they can go to the assistant superintendent who, again, will direct-- call back to the school and say, "Remedy the situation. Do something about it." Sometimes we have to tell parents things they don't want to hear, but I do think we go out of our way to accommodate parents. . . the reason we do that is not for fear they're going to drop out of school because we don't think that's going to happen, but I think it's because of an attitude in this district that says that parent calls are very, very important. . . I think that the tone that the school board even sets. They have these open microphones at every board meeting . . . . the superintendent will receive a call, for example, and she will personally call the school and ask what the situation is. I mean, one call in a district as large as this means a lot and that's just the attitude of this district.

Return to the Table of Contents

Teacher and Principal Constraints, As They Describe Them

Private and public schools are subject to many of the same constraints. Constraints appeared in the form of requirements imposed by college admissions and the College Board, financial pressures, the threat of litigation, and parent demands.

College admissions requirements force prospective students to take specified courses. The College Board, through advanced placement testing, delineates a specified curriculum in specialized subject areas so that students taking the test will be successful. Passing the test confirms that the student has fulfilled the curriculum equivalent to an entry-level college course. The same admissions requirements and the same advanced placement tests apply to all secondary schools regardless of their organization or distinction as public or private. Teachers and administrators alike pay a great deal of attention to this area because both parents who send children to college preparatory independent schools and parents of college-bound public school students expect their children to take the courses required by the better colleges. The stakes, therefore, are highest in this area. Little constraint, however, is felt by teachers, whether public or private, of non-college preparatory courses. Parents, and therefore administrators, pay little attention to these courses, thus permitting these teachers considerable freedom.

Financial constraints limit options available to any type of school and its teachers. In both the public and private schools, finances often determine class loads and class size. The availability of many instructional resources is largely determined by available funds. Private schools spend considerable time and effort soliciting additional funds for these purposes. Indeed, the primary function of the board of a private school is one of raising funds and, for the schools in this study, establishing an endowment fund. The public schools rely on support from the community at large when requesting additional funds through bond elections.

The threat of litigation affects public and private schools equally. Civil laws and laws regarding the health, safety, and welfare of students and employees do not differentiate between public and private institutions. Since schools of any organizational structure are equally susceptible to litigation based on the same set of laws, all schools experience this constraint to the same degree. It is acknowledged, however, that some laws pertain to public schools and exempt private schools since private schools are able to avoid students with special needs. Public schools serve the needs of all students and are obliged to provide equitable services. Additional federal and state mandates require public schools to function in a bureaucracy at least large enough to handle their administration (Boyer, 1983, p. 226).

Chubb and Moe (1990) painted a picture of public school educators oblivious to the opinions and wishes of parents, tending a bureaucratic institution that has lost touch with its clients. These were not the educators who spoke of the pressures they felt to meet parents' expectations. There is no lack of concern for the expectations of parents of children who attend public schools. Parents of college- bound students are often highly educated professionals who are vocal in making demands on the schools and teachers. Grant (1988) reported that "in the aggregate parents as a whole may now be more educated relative to teachers and thus are likely to be more critical of teacher performance" (p. 149). Whether public or private, teachers and principals reported frequent contact with parents, making parents feel part of the team or family, and sharing the same expectations as parents for the children. In the public schools, that response at times included a response from the school board. Parents in large public school districts use the bureaucratic layers as alternative audiences to make their voices heard.

Autonomy in Public and Private Schools

The feelings expressed by all of the participants in this study, both private and public, testify to a high degree of autonomy. The responses to interview and survey questions alike clearly dispel the myth that autonomy is generally high in private schools and generally low in public schools (Chubb & Moe, 1990, p. 183). Autonomy is generally high in both types of school studied here. Issues that emerged in the course of this study from teacher and administrator descriptions of their autonomy are: conflicting and contradictory demands, shared beliefs, layers of protection, a system of laws, funding constraints, and matters of size of the institution. These issues challenge oversimplified assertions that differences of any significance exist between the perception of autonomy held by professionals in public or private high schools.

Before embarking on a detailed examination of the concept of educator autonomy, it is well to emphasize the particular characteristics of the sites examined here and how that characteristic may shape what has been learned. The educators who consented to be interviewed practice in upper-middle class college preparatory public and independent, non-public secondary schools. One might not expect to learn the same things about autonomy in religious affiliated private schools, though it is unclear whether the autonomy would be expected to be greater or less. One must also be cautioned about extrapolating the insights garnered from this research to other levels of school, such as elementary.

The schools examined here enjoy success in all conventional senses of the term. This favorable environment may shape the way the political system treats educators and how educators respond in return. One might have reason not to expect the same organizational effect obtaining in schools under the duress of poverty and social dislocation. The following themes that emerged from this research should be viewed with these cautions in mind.

Conflicting and Contradictory Demands

Contrary to the popular myth of public school bureaucratic insulation and insensitivity, both the public and private high schools in this study showed a sensitivity and prompt response to parent concerns. Parents are listened to and given serious consideration. Parent and teacher communication are encouraged in both the public and private high schools. Parents have access to the administration as well as teachers. In the larger system of the public schools, parents receive additional attention from the superintendent and school board. In all cases, board meetings are open to parents with one public system, in particular, scheduling an "open mike" segment prior to handling business on the agenda. Another public school conducted a survey of parent expectations that determined the goals of the school.

Being responsive to parents has the potential, however, of constraining the very autonomy that some deem a requirement for creative and innovative teaching. The principal or head of school prevents responsiveness to parents from becoming a constraint on teacher innovation by virtue of a strong belief about how students are best served. The criterion which defines the degree of autonomy granted is based on the perceived success of the school and its students. But underlying the freedom of teachers and principals is a clear understanding of what parents and the community expect of the schools. If parents perceive the school to be doing what they say they are doing, public or private school teachers and administrators experienced greater freedom and fewer external pressures.

Being responsive to parents prompted an assistant head of a private school to claim they must work "on the conservative side." It is what led one public school teacher to admit, despite opportunities to have a voice in decisions that affect her work life, that she felt little freedom and great frustration knowing those decisions must be responsive to a conservative parent body. This unspoken tension between autonomy and obligation requires teachers and administrators in both public and private schools to "negotiate competing demands" (Hawthorne, 1988, p. 231). Hawthorne's study found this negotiation process for the teachers she studied to be as individual as the negotiators (p. 231). In this study the manner in which teachers balance the demands of parents, administrators, or others who attempt to influence them with their own need for control was, similarly, an individual matter. Some chose to ignore certain rules or policies which they deemed insignificant, others relied on the support of their principal, head, or department chair.

Both public and private school teachers and their administration demonstrated a responsiveness to parents by focusing on the needs of college-bound students. A college preparatory independent school and public high school housing a student body for whom college admission is a high priority are forced to meet the requirements of those colleges and the college advanced placement program. The curriculum of AP courses offered in both the public and private high schools are extensively defined by the College Board, which administers the exams students take to earn college credit. The advantage of size of the public institutions is that they can often offer a larger number of AP courses. Their size necessitates the offering of a number of the same courses to meet the demands of those students who qualify. Teachers of non-college preparatory courses describe the greatest degree of autonomy in curriculum decisions and all areas of teaching. An acknowledged lack of parent interest gives them this heightened sense of freedom.

Parents of college-bound students are vocal in the public schools. The voice of parents of the college-bound student is heard and heeded. The demands and pressures placed on the school are felt to an equal degree in the private schools. Both types of schools must balance autonomy and obligation to parents.

The private school response must include a consideration of consequences to the institution. If parents are not satisfied, the funds upon which the school depends can be withdrawn. The existence of the private school depends on satisfying the parent community. Even within the college preparatory private school, however, there exists a range of demands to which the head must respond. Those who imagine that private schools are very responsive to "customer" (or parent) demands or needs overlook one significant fact about American education: even small, homogeneous publics make conflicting and often contradictory demands. How is the school supposed to accommodate these wishes when one faction calls for greater emphasis on algebra and another calls for less? All three of the private schools in this study are small, with a constituency composed of middle-class to high income families and a desire for the kind of academic program which will enhance college admission. Even within this situation, heads of these schools found themselves taking a stand. Could they really afford to finance an advanced calculus program despite the demands of a few parents? Should a teacher whose personality was not tolerated by some parents be fired?

It is because of conflicting parent demands within the small private school setting that a head of school declared, " It is our responsibility to be service-oriented and to be responsive to our parents, but it is not our responsibility to place them in the position of calling the shots." Not every parent can get his or her way. The head must take a stand to protect the autonomy of the teacher. Although the public schools in this study were selected because they, too, focused on academic preparation for college, they were likewise not able to escape contradictory demands of parents. One principal stated, " . . .they [parents] impact me on a daily basis, but they don't drive me."

Principals in public schools are not necessarily threatened by withdrawal of a student, but they are under pressure by the community at large to respond to the needs of the students and demands of their parents. The bureaucracy works to the parents' benefit. If satisfactory recourse is not forthcoming from a teacher or principal, parents may voice their concerns to a superintendent or school board which has been elected to represent them. Educated and politically active parents know how to get things done despite a large bureaucracy. In turn, these layers of administration can preserve teacher and principal autonomy by providing support or acting as a buffer.

Shared Beliefs

Teachers in independent schools talked about the freedom they have to design their own curricula, utilize any variety of teaching methods, and select their own textbooks. Coordination among faculty within a department or between a lower and upper school is often informal if it occurs at all. They share basic educational beliefs. These teachers were, for all practical purposes, hand-picked by the head for just that reason. The head of an independent school can comfortably allocate substantial portions of autonomy to these teachers.

Principals in public schools also report considerable freedom in selection of faculty. They, too, choose teachers who share the same philosophy and an understanding of the expectations of the parents. It is with confidence, they can trust teachers to make appropriate decisions and provide the autonomy that teachers experienced. As long as the school runs smoothly, there is little need to question teacher autonomy. Indeed, the goal of site-based management within individuals schools assumes teachers will make the kind of decisions that support parent expectations and the goals of the school. In site-based schools teams of same subject teachers hire new faculty. There are, admittedly, times when principal freedom to select teachers is curtailed. This occurs when the relocation of a teacher from another high school within a district is due to a reduction in student population, therefore, a reduction in faculty. In this case, a principal would be required to take a particular teacher.

The concept of shared beliefs of teachers and principals, schools and the community is the issue from which the perception of autonomy stems. In some sense, the autonomy that teachers, whether public or private, feel in relation to their principals is like the proverbial equity allocated to both beggars and rich men:

> The law in its majestic equality, forbids the rich as well as the poor to sleep under bridges, to beg in the streets, and to steal bread. (Anatole France, Le Lys Rouge)

What sort of freedom is it if it is never tested by conflict? When a principal selects teachers who generally agree with the principal's beliefs and values, there will be few conflicts and few instances when the teachers have to be told what to do by the principal. Some would say, then, that the principal has not, in fact, granted autonomy to the teachers, or that the limits of their autonomy are untested. They seldom chose to do that which would be overturned by their principal.

Much of the autonomy that teachers feel may be of this type. If so, teachers might be better described as "like- minded" with their superiors rather than autonomous in relation to them. In either case, however, the image created by this view of like-minded or autonomous teachers is quite different from the image drawn by some in which teachers are portrayed as deadened and oppressed by a hierarchical bureaucracy. The challenge of making schools creative, interesting and productive environments for students may be more a matter of stimulating teachers and principals who have fallen into complacency than to free them from some ill-conceived notion of an repressive and domineering bureaucracy. It is important to recognize, however, that the above situation could be quite different in elementary schools or in secondary schools suffering the effects of under-investment and the pressures exerted by special social problems.

Layers of Protection

Unlike the traditional perception of public high school bureaucracy, the hierarchy that exists is built out of a necessity to manage large numbers of people and a complex institution. The teachers have a number of layers to protect them from external influences. The department head is one line of defense and a person who speaks on behalf of the faculty of that department. The principal and assistant principal also protect teacher autonomy. Equally strong is the sense of control teachers feel because of the security that the teachers' association provides. The superintendent can also be a buffer between the principal and teachers and the school board. The trust that principals and department chairs expressed in their faculty is not unlike that described by the heads of school in this study. Acting as a buffer, however, does not mean to ignore the wishes of the parents.

While public and private school teachers have the advantage of protection from administrators who demonstrate support in their professionalism, public school teachers have access to an additional entity. Membership in teachers associations provides another layer of protection for private school teachers. Those who claim that teacher autonomy is a requisite of the best education should applaud the teachers associations for giving teachers the kind of security they need to feel truly autonomous. This protection, in many cases, gives teachers the sense of control they need to try out innovative or creative ideas. Knowledgeable principals and teachers in public schools are able to use the teachers association to preserve their autonomy.

A System of Laws

Freedom in both public and private high schools is constrained to an equal degree by a system of laws. These laws protect the health and safety of the inhabitants of both types of school. All schools have fire drills and public safety requirements. Health issues are promptly dealt with by both public and private institutions. Worker rights are addressed in the private schools by the same type of union that protects workers in any institution. Civil laws protect the basic rights of teachers and students. The threat of lawsuits has an equal effect on both public and private schools and influences many decisions made by teachers and administrators.

Funding Constraints

A lack of an appropriate level of funding is yet another constraint on teachers, principals, and heads of school alike. Private schools cannot offer the range of courses offered in the public schools because the small numbers of students enrolled in each class will not support the cost of an additional teacher. Decisions constrained by finances result in large class size in the public school. The availability of certain instructional materials, such as computers, is often determined by finances rather than choice. Financial constraints put limits on the autonomy of both public and private schools.

A Matter of Size

The three private schools in this study are small in student population, faculty, and facility. The public schools by comparison are larger in each category and require a degree of bureaucracy to manage the sheer size. The size also necessitates articulation of curriculum among grades and a means for frequent communication among groups of teachers. Teachers in the public schools are encouraged to participate individually or through their representatives in policy making. Bureaucracy may make greater demands of teachers' time to participate in decisions that affect them. But it is apparent that it does not impinge on teacher freedom over those decisions that matter most to them--the decisions that affect what occurs in the classroom.

Because of the size of the public institution, teachers rely more on their colleagues and department chair both for advice on decisions that affect them and for protection from external influences. They have opportunities for participation in decision-making through committee work, access to department chairs, and access to the principal. Teachers in private schools express autonomy in similar ways. They have less of a need for a department chair to be a spokesperson merely because of the proximity of the head. The size of an organization cannot be ignored; however, but neither can it be called a determinant of teacher autonomy.

Little of the quality of what occurs in the classroom can be defined by the size of the institution. There is a general belief that private schools equate with academic excellence. The perception of academic excellence in private schools may stem from a belief that small schools are less complex and small classes necessarily produce a quality education. Large public schools do not offer small classes, but they can offer college preparatory courses and advanced placement classes. What occurs within each classroom is under the control of individual teachers whether public or private, as the teachers in this study have described.

When Chubb and Moe described "ineffective" and "effective" schools, they were essentially referring to public and private schools, respectively. Their critique of the organizational structure of each does not involve a comparison of organizational units of the same size or of like populations between the two types of schools. As reported by Hogan (1992b), since "public and private schools are very different kinds of schools that recruit different populations, pursue different objectives and tasks, and develop different tools to achieve them, comparing them is like comparing apples and oranges" (p. 93).

Conclusion: The Myth of the Market

The findings of this research challenge directly the assertions made in one of the most visible research documents on the question of school choice, viz., Chubb and Moe's *Politics, Markets and America's Schools.* (1990). They acquired their data from the High School and Beyond Survey (Moles, 1988) and performed a secondary analysis of this government survey to collect information on effective schooling. Several critics argued that there were weaknesses in their analyses and their interpretation: Witte (1992), Hogan (1992a, 1992b), Goldstein (1992), and Glass & Matthews (1991). Some said that they were unable in their book and in their analysis to determine whether it was effective schools that were granting autonomy to their teachers and administrators or whether autonomous teachers and administrators were producing more effective schooling (Glass & Matthews, 1991). In other words, the direction of the influence may be reversed; that it may be the perception of a successful school (advanced placement courses, National Merit scholars, high graduation rate, admission to elite colleges) that confers autonomy to teachers and administrators. In addition, even though in their book, they tried to argue that private schools would necessarily grant more autonomy to teachers and administrators than public schools, Chubb and Moe never once analyzed or reported data from the High School and Beyond Survey on that question. In fact, they presented no data whatsoever from private schools, claiming that the data base in the survey was inadequate for making any generalizations. However, that did not stop them from making claims about the superiority of private schools (and, hence, the superiority of choice as a policy) because they assumed private schools grant more autonomy and demonstrate more responsiveness to parents and market pressures than public schools. Among the assertions made by Chubb and Moe, three are directly refuted by the findings of this research study: markets, bureaucracy, and the role of teachers unions or associations.

On Markets

Chubb and Moe sought to perpetuate the myth that only private sector schools experience the goading of the market- place. They wrote:

> Under a system of democratic control, the public schools are governed by an enormous, far-flung constituency in which the interests of parents and students carry no special status or weight. When the markets prevail, parents and students are thrust onto center stage, along with the owners and staff of schools; most of the rest of society plays a distinctly secondary role, limited for the most part to setting the

> framework within which educational choices get made (p. 35).

If responding to market pressures means responding to parent demands, the public schools are doing just that. To a very substantial degree, market pressures of various kinds have shaped and continue to shape educational politics and the institutions affected by them. The relatively decentralized structure of educational politics in this country actually enhances the vulnerability of school officials to popular political pressures, and, thus, to the market forces that shape educational politics. Parents are not without choice, or voice. Hogan (1992b) pointed out, "Savvy school officials . . . respond to the underlying anxieties and aspirations [of parents] by rigorously tracking . . . or by creating magnet schools or in any number of ways--parent choice being the latest--to keep their middle-class constituency from fleeing the public schools" (p. 193). The parents to whom he refers are those who believe college is the route to attain or maintain a middle to upper social class standing that they want for their children. The admittance of students to what Powell and others (1986) call top-track "specialty shops" (p. 124) forms the basis of the willingness of parents to enroll their children in public school system. These are not unlike the communities to which the public schools in this study belong. Student achievement within such schools is a matter of residential pattern, social demography, patterns of political participation from members of the community, and leadership in local educational politics. Parents whose children attend private schools may not share the same residential community, but they do share social demographics as well as participation and leadership in the workings of the independent school. In this study, teachers and administrators in both the public and private high schools supported Hogan's contention that they are responsive to parent expectations. College preparatory courses were given a great deal of attention and advanced placement courses were instituted. The public schools use advanced placement and upper level content courses as a tracked curriculum and become like a private school within a public school in response to parent demands.

On Bureaucracy

Chubb & Moe decry the oppression of bureaucracy in the public schools and commend the private schools for their lack of bureaucracy, therefore, creating greater autonomy than possible in the public school. They claimed:

> . . . we show that private schools are organized more effectively than public schools and that this is a reflection of their far greater autonomy from external(bureaucratic) control (p. 24).

Chubb & Moe further stated:

> Its [public school] institutions of democratic control are inherently destructive of school autonomy and inherently conducive to bureaucracy (p.47).

Teachers in both kinds of institution reported feelings of considerable autonomy in such matters as determination of curriculum, dealing with students, parents, curriculum development. Not only was there a strong statement of autonomy on the part of these teachers, it was impossible to distinguish any difference in the strength of those feelings between public and private institutions. Additionally, two questions were also analyzed from the High School and Beyond Survey that Chubb & Moe had given and yet never used to report a comparison of public and private school teachers. When analyzed to compare teacher ratings of autonomy in several areas of their functioning (determining student behavior codes, content of inservice, curriculum, policies of grouping students, textbook selection, choice of teaching methods, etc.) the average scores came out virtually equal between public and private which increased suspicions that perhaps Chubb and Moe had seen in the High School and Beyond Survey no differences between public and private and were disappointed or confused by it. Perhaps it did not agree with their expectations about markets and choice in school so they chose not to report it.

Many public school teachers in this study reported that the bureaucracy was supportive and protected their autonomy. One teacher spoke for many others when she said that the principal, superintendent, and school board did not give in "to cranky parents." This trust in the professionalism of the teacher gave many the perception of autonomy. What may appear to be a contradictory notion is the idea that knowledgeable parents understand the large public school system and are able to make it work for them. If they are dissatisfied with the response of a particular teacher, they can find a voice with the principal, superintendent or school board. Opportunities to be heard are found at each layer of the hierarchy. It is possible for the bureaucracy to be responsive to parents demands, yet make teachers feel they are not subject to the whim of the occasional "cranky parent."

What of the question of shared beliefs? If principals or heads select and hire teachers who are like-minded, is autonomy really tested? In this case, teachers are more appropriately termed like-minded rather than autonomous. Teachers who share an education philosophy with the administrator can be trusted, given support and wide latitude leading to a perception of autonomy. In a similar vein, teachers reported that as long as things were going smoothly, no parental complaints, they felt greater freedom. These ideas were found to a strikingly similar degree in both the public and private high schools. In any case, regardless of the source of the perception of autonomy, in no instance were teachers perceived to be oppressed or deadened by the weight of bureaucracy. The challenge of making schools more creative, energetic, and innovative institutions may more be a matter of stimulating teachers and principals who have fallen into complacency rather than setting them free from some ill-conceived notion of a repressive and domineering bureaucracy.

On Teachers Associations

The role of the teachers union in constraining the autonomy of teacher is described throughout the text of "Politics, Markets and America's Schools." (1990). This study was completed in a right-to-work state, where public school teachers have come to look to teachers associations rather than unions to protect their working conditions. For all intents and purposes, these teachers associations and teachers unions are synonymous. Chubb and Moe claimed:

> Teachers who are team players, who have lots of autonomy in their work, who routinely play integral roles in school decision-making, and who are treated as professionals are hardly good candidates for union membership (p.53).

The public school teachers in this study reported that the teachers association actually protected their autonomy. For example, parents cannot force a teacher to change a student's grade. It is also the teachers association which negotiated a contract requiring teacher participation on committees charged with making policy decisions. Even principals spoke of working with the teachers association on controversial matters. The position of the association helped administrators by providing clear guidelines, thus avoiding ambiguity on many issues including the firing of teachers. Both administrators and teachers claimed the association's role was to guarantee fairness in workplace conditions. The public secondary school teachers in this study all reported opportunities to participate in decisions that are important to them, contrary to the blanket statements put forth by Chubb and Moe.

Administrators in the public school reported considerable autonomy in the hiring and firing of teachers. It is true they work within the confines of the teachers association to fire teachers, however, their ability to fire teachers is not thus impaired. Experienced administrators understand how the system works and do not feel constrained. Two public school principals told of having freedom in the hiring of personnel. On rare occasion a principal may be required to accept a teacher who has been transferred due to a reduction in work force at another of the district's schools. The third principal, who reported reduced autonomy in the hiring of teachers, described how the faculty and chair of each department has the primary responsibility for hiring new faculty to their department. He was unconcerned about their selection because he trusted them. Teachers who share the educational philosophy of the administrator can be expected to hire new faculty with similar beliefs. This principal has an indirect role in the hiring of teachers. In this study, none of the public school principals experienced constraints in hiring or firing faculty as was the experience of their private school counterparts. These perceptions of autonomy exist despite the presence of teachers associations

Conclusion

Chubb and Moe were perpetuating a general view regarding public and private schools. They sought to perpetuate the myth that teachers and principals in private schools enjoy autonomy and freedom from democratic bureaucracy that their public school counterparts do not. They further claimed that private schools only are subject to market forces.

This research shows how complex the reality is. Autonomy is an issue that does not clearly distinguish public from private education. The freedom teachers and administrators feel and the constraints they experience are complex. Many of the constraints experienced by public and private high school administrators and teachers are similar. Both sectors must work within the limits of a set of prescribed laws. They are equally subject to pressures resulting from limited funds. Perceptions of autonomy are individual matters, often experienced within a range of accepted constraints. Teachers and administrators describe their attempts to secure professional autonomy in an arena circumscribed by the demands of parents, college admissions requirements, and the College Board. Often these demands are conflicting and contradictory, yet teachers are able to exert autonomy by seeking protection from administrative hierarchies, participating in opportunities for decision- making, ignoring selected policies, and seeking the sanctuary of their own classroom where their authority is unchecked. The greatest freedom is derived from the perception of a successful school. In schools that produce students who gain admission to colleges of choice, as in this study, teachers, principals and heads of school enjoy considerable autonomy.

The limited scope of this study points to the need to explore perceptions of autonomy in the context of other types of private schools. How do teachers and administrators in religious private schools experience autonomy? In what ways do the social and economic circumstances of the students affect teachers' and administrators' autonomy? What are the relationships between autonomy and achievement when the variables of religious affiliation and economic level differ from those in the present study? Given the complexity of the issues and the persistence of the debate about privitization of education further research on autonomy is warranted.

Return to the Table of Contents

References

Apple, M. W., & Teitelbaum, K. (1986). Are teachers losing control of their skills and curriculum? *Journal of Curriculum Studies*, 18, 177-184.

Baird, L. L. (1977). *The elite schools: A profile of prestigious independent schools.* Lexington, MA: Lexington Books.

Ball, S. J. (1987). *The micro-politics of the school.* New York, NY: Methuen and Company.

Bierlein, L., Sheane, K., & McCarthy, K. (Eds.) (1991). Open enrollment/Educational choice. Unpublished paper. Tempe, AZ: Morrison Institute for Public Policy, Arizona State University.

Blase, J. J. (1988). Dimensions of ineffective school leadership: The teacher's perspective. *The Journal of Educational Administration*, 24, 193-213.

Blase, J. J. (1991). Some negative effects of principals' control-oriented and protective political behavior. *American Educational Research Journal*, 27, 727-753.

Boyer, E. L. (1983). High school: A report on secondary education in America. The Carnegie Foundation for the Advancement of Teaching. New York: Harper & Row, Publishers.

Chubb, J. E. & Moe, T. M. (1985, August). Politics, markets, and the organization of schools. Paper presented at the Annual Meeting of the American Political Science Association, New Orleans, LA.

Chubb, J. E. & Moe, T. M. (1990). *Politics, markets, and America's schools.* Washington, DC: The Brookings Institution.

Coleman, J. S. & Hoffer, T. (1987). *Public and private high schools: The impact of communities.* New York, NY: Basic Books.

Corbett, H. D. (1991). Community influence and school micropolitics: A case example. In J. Blase (Ed.), *The politics of life in schools* (pp. 73-95). Newbury Park, CA: SAGE Publications.

Elmore, R. F. (1987). Reform and the culture of authority in schools. *Educational Administration Quarterly*, 23, 60-78.

Feiman-Nemser, S. & Floden, R. E. (1986). The cultures of teaching. In M. C. Wittrock (Ed.), *Handbook of research on teaching.* (Third edition). New York, NY: Macmillan Publishing Company.

Firestone, W. A. & Bader, B. D. (1991). Professionalism or bureaucracy? Redesigning teaching. *Educational Evaluation and Policy Analysis*, 13, 67-86.

Freire, P. (1981). *Pedagogy of the oppressed.* New York, NY: Continuum.

Glaser, B. (1978). *Theoretical sensitivity: Advances in the methodology of grounded theory.* Mill

Valley, CA: Sociology Press.

Glass, G. V. & Matthews, D. A. (1991). Are data enough? Review of *Politics, markets, and America's schools. Educational Researcher*, 20, 24-27.

Goldstein, H. (1992). Review of "Politics, markets, and America's schools." Posted on EDPOLYAN, a LISTSERV at ASUVM.EDU, on November 29, 1992. 241 lines. (Retrievable by sending the e-mail message GET EDPOLYAN LOG9211 to LISTSERV@ASU.EDU.)

Grant, G. (1988). *The world we created at Hamilton High*. Cambridge, MA: Harvard University Press.

Hawthorne, R. K. (1988). Classroom curriculum: Educational criticisms of teacher choice. (Doctoral dissertation, Stanford University, 1988). Dissertation Abstracts International, 49, 08A.

Hogan, D. (1992a). School organization and student achievement: A review essay. *Educational Theory*, 42, 83-105.

Hogan, D. (1992b). ". . . the silent compulsons of edonomic relations": Markets and the demand for education. *Educational Policy*, 6, 180-205.

Kozol, J. (1981). *On being a teacher*. New York, NY: The Continuum Publishing Company.

Lightfoot, S. L. (1983). *The good high school: Portraits of character and culture*. New York, NY: Basic Books.

Lortie, D. (1975). *The school teacher: A sociological study*. Chicago, IL: University of Chicago Press.

Mathison, S. (1988). Why triangulate? *Educational Researcher*, 17, 13-17.

McNeil, L. M. (1986). *Contradictions of control: School structure and school knowledge*. New York, NY: Routledge.

Miles, M. B. & Huberman, A. M. (1984). *Qualitative data analysis: A sourcebook of new methods*. Beverly Hills, CA: SAGE Publications.

Moles, O. (Ed.) (1988). High school and beyond administrator and teacher survey (1984): Data file users manual. Washington, D.C.: Office of Educational Research and Improvement, U. S. Department of Education.

Noble, A.J. & Smith, M.L. (1994). Old and new beliefs about measurement-driven reform: Build it and they will come. *Educational Policy*, 8 (2), 111-136.

North Central Association of Colleges and Schools. (1990). Policies and standards for college preparatory schools. Tempe, AZ: Commission on Schools.

Popkewitz, T. S. & Lind, K. (1988). Teacher incentives as reform: Teachers' work and the changing control mechanism in education. *Teachers College Record*, 90, 575-594.

Powell, A. G., Farrar, E., & Cohen, D. (1986). *The shopping mall high school*. Boston, MA:

Houghton Mifflin.

Powell, A. G. (1990). The condition of teachers' work in independent schools. Pp. 111-141 in M. W. McLaughlin, J. E. Talbert, & N. Bascia (Eds.), *The contexts of teaching in secondary schools: Teachers' realities.* New York, NY: Teachers College Press.

Russo, C. J. (1990). Review of Kerchner, C.T. & Mitchell, D. E. *The Changing Idea of a Teachers Union. Education Administration Quarterly*, 26, 192-195.

Sedlak, M. W., Wheeler, C. W., Pullin, D. C. & Cusick, P. A. (1986). *Selling student short: Classroom bargains and academic reform in the American high school.* New York, NY: Teachers College Press.

Smith, M.L. & Rottenberg, C. (1991). Unintended consequences of external testing in elementary schools. *Educational Measurement: Issues and Practice*, 10, 7-11.

Staff. (1991, November 12). Education reform task force: Symington pulls out the rug. "The Arizona Republic", p. A10.

Witte, J. F. (1992). Public subsidies for private schools: What we know and how to proceed. *Educational Policy*, 6, 206-227.

Wolcott, H. F. (1990). *Writing up qualitative research.* Newbury Park, CA: SAGE Publications.

Return to the Table of Contents

This article has been retrieved 4216 times since January 8, 1997.

Education Policy Analysis Archives

Volume 5 Number 2 January 8, 1997 ISSN 1068-2341

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass,Glass@ASU.EDU. College of Education, Arizona State University,Tempe AZ 85287-2411

Copyright 1997, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

Where Have All the Teachers Gone?

Mark Fetler
California Commission on Teacher Credentialing

Abstract

A rising need for teachers is projected for California and the nation during the next decade. Sound policy for teacher preparation should not only foster a capable workforce, it should also assure that the supply of qualified teachers balances with employment demand. A conceptual model is proposed to describe the flow of individuals through teacher preparation programs and the workplace. In California the workforce is projected to grow by thirty percent over the next ten years, stimulating the demand for teachers. At present the number of newly credentialed teachers exceeds the number hired. However, the apparent abundance masks an oversupply of teachers in some curricular and geographic areas and shortages elsewhere. Evidence for a lack of balance between supply and demand is found in an upward trend of emergency hiring of teachers who do not meet all requirements for a credential and low employment rates for first-time college and university prepared teachers. The asymmetry between supply and demand could be redressed partly through better retention of working teachers and closer coordination of preparation programs with the needs of schools in their service areas.

A basic concern for state and federal policymakers who fund and regulate public school systems is determining how many teachers are needed to provide a desired level of service to a given student population. The number of students assigned to a teacher, a measure of workload, presumably influences the way in which a teacher prepares, delivers instruction, manages the classroom, etc. Conventional wisdom suggests that lower teacher workloads should result in more attention to individual students, and stronger student outcomes. The Tennessee studies of class size found that teachers in smaller classes have more time to give to individual children. (Mosteller, Light, and Sachs, 1996) In Tennessee a reduction in class size from 23 to 15 in grades K 3 speeded up learning and continued to confer lasting benefits to students when they attended larger classes in later grades. Even so, there is ongoing debate among researchers about the relationship between class size and student achievement. (Glass, 1979; Greenwald, Hedges, and Laine, 1996; and Hanushek, 1996)

Such debate is understandable given the increased costs of smaller classes. For example, California policymakers in 1996 allocated $771 million for a statewide reduction in class sizes for grades K-3. According to Kirst, Hayward, and Koppich (1995) California appropriations for K-12 education consume 35% of the state's general fund, with teacher salaries accounting for 80 percent. The average annual teacher salary in California is about $40,000. In 1995-96 there were about 5.4 million students enrolled in California's public schools, along with 232,000 teachers, yielding a student-teacher ratio of 23:1. In order to reduce this ratio by one point to 22:1, it would be necessary to increase the pool of working teachers by 10,000 at a projected cost of $400 million. Of course, actual costs could be more or less depending on how the increase is achieved. The pool of working teachers could be enlarged by slowing the transition of teachers out of school employment, by recruiting more formerly employed teachers back into service, or by recruiting more first time teachers from traditional or non-traditional preparation programs. Whatever the methods, employing more teachers, means drawing money away from competing policy goals, a decision which is attended by debate.

Teacher Supply and Demand

An evaluation of the supply of teachers in relation to demand provides relevant background for such policy choices. There are at least two significant dimensions to an analysis of supply and demand. One dimension relates to the skills and abilities expected of teachers. For example, California has a linguistically and culturally diverse student population. Many of California's teacher preparation programs have added training to facilitate adapting instruction in culturally appropriate ways and language acquisition. While more extensive teacher preparation is intended to improve instruction, it usually consumes additional time and resources. Darling-Hammond and Hudson (1990) and the report from the National Commission on Teaching and America's Future (1996) discuss teacher preparation issues in detail. Reynolds (1991) comments that teacher licensure is intended to protect the public from harm. Ashton (1996) notes that teachers with regular state certification receive higher supervisor ratings and student achievement than teachers who do not meet standards. If teachers are not adequately prepared for their jobs and cannot teach effectively, they place their students at risk. The requirements for a teaching credential are designed to provide assurance that teachers are adequately prepared. Ideally, such requirements reflect a consensus of expert practitioners, teacher educators, and researchers. The substance of teacher preparation is not the main focus of this paper.

The second dimension of an analysis of teacher supply and demand describes the flow of people into and out of public school employment. Primary components of this dimension are the hiring needs of public schools and the capacity of various sources to meet those needs. Major sources of credentialed teachers include college and university preparation programs and re-entrants from the reserve pool of previously employed teachers. Other numerically less prominent sources are school district and university programs to facilitate the mid-career transition of people into teaching from jobs in other industries or the military.

The relationship between the two dimensions of preparation and flow is complex. One hypothesis is that certification and licensure requirements restrict access to the teaching profession. Other conditions remaining equal, higher standards will depress the numbers of teachers who are prepared. Under this hypothesis there are two ways to meet increased demand.

One way is to lower the requirements, reducing the time and cost required to become a teacher. A risk of this strategy is that less well prepared teachers may tend to be be less effective in their jobs and more prone to attrition. A second way is to provide additional incentives to prospective teachers, for example, higher salaries or better working conditions.

When there are insufficient numbers of suitably credentialed applicants, California school districts can hire individuals on emergency permits who lack some requirements for a credential, usually proof of competence in their subject(s) of instruction or pedagogy. (Hart and Burr, 1996) Approximately one-third of emergency permits in 1994-95 were issued to individuals without the training to teach or work with children. Emergency permits are most often granted in the areas of special education, bilingual education, mathematics, and science. Emergency permit hiring is more prevalent in California's large urban districts than elsewhere. Where the need is ongoing, emergency permits are renewed annually. About sixty percent of California's emergency teachers are teaching on renewals of old permits, with no limit on the number of possible renewals. Overall the percentage of teachers on emergency permits has grown steadily since 1989, with much but not all of the increase related to more stringent credentialing requirements in special education.

School district demand for teachers is influenced by the willingness of prospects to apply for jobs. Tierney (1993) surveyed employment decisions of recent graduates of teacher preparation programs and school district personnel offices in California. Most graduates intended to teach at public schools within California and began searching for employment during or immediately following the completion of their programs. Over half preferred to teach within 25 miles of their current home. The most important reasons for applying to their chosen districts included: closeness to their current home, availability of assignment, and reputation of the district. The school district priorities for evaluating candidates included: job interview, performance in student teaching, candidate enthusiasm, and reference letters. Although school districts generally require transcripts, the reputation of the credential program, the candidate's academic record, and the reputation of the undergraduate institution were less heavily weighted.

An additional factor influencing the demand for qualified teachers is the rate at which teachers leave public school employment. One challenge in estimating teacher attrition is that few research studies continue long enough to observe the departure of all participants. Willet and Singer (1991, p. 411) comment that "the clearest signals about teachers' careers will come from studying cohorts of teachers whose professional lives were tracked from a common reference point - their entry into teaching." They recommend survival analysis as a statistical method for coping with such "censored" data. Longitudinal tracking of individuals permits the calculation of various statistics, including survival probabilities, or the proportion of an initial cohort surviving through successive years. Median career length is computed as the elapsed time until half the cohort has left. Hazard probabilities are the conditional probabilities that a teacher will leave, given that he or she survived through the end of the previous year.

Not all teachers who quit permanently sever their ties with public schools. A significant part of the demand for teachers can be met by rehiring individuals who have previously taught. Beaudin's (1993, 1995) studies describe factors influencing the reentrance of previously employed Michigan teachers into public school service. Teachers were more likely to return if their instructional specialties provided limited opportunities for higher paid jobs outside of education. They were more likely to return if they had more than two years of teaching experience coupled with a masters degree, or if they were older when they interrupted their career. About 55 percent of reentering teachers returned to their original school districts. The probabilities of returning to the original district were higher for those districts with higher salaries and higher levels of funding. Individuals with more years of teaching experience and who only interrupted their careers for one year were more likely to return to their original districts.

Nationally, between 1988 and 1994, schools hired increasingly larger proportions of first-time teachers and smaller proportions of reentrants. (NCES, 1996a) In 1994 about 57 percent of first-time teachers came fully prepared from college or university programs, a decrease of about 10 percentage points since 1988. Possibly a consequence of this decrease, many students are being taught core academic subjects by teachers without adequate educational qualifications in their assigned fields. (NCES, 1996b) For example, in grades 7-12 during 1990-91 about one fifth of students received instruction from underqualified teachers in English, one-quarter of students in mathematics, thirty-nine percent in life science or biology, fifty-six percent in physical science, and over half in history or world civilization.

NCES (1996c) estimates that total K-12 enrollment will increase from 49.8 million in 1994 to 54.6 million by 2006, an increase of about 10 percent. During the same period the number of high school graduates is estimated to increase by 21 percent. Perhaps reflecting uncertain prospects for growth in higher education, NCES projects the number of bachelor's degrees to increase, either by 0.5 percent under a low alternative, or by 22 percent under a high alternative. Under a middle alternative, the number of classroom teachers is expected to increase from 2.96 million in 1994 to 3.43 million by 2006, a rise of 16 percent.

Supply and Demand Model and Indicators

An evaluation of teacher supply and demand should include estimates of the numbers of teachers needed, along with the capacity of primary sources to meet that demand. Ideally, a flow analysis would track individuals as they make their way through postsecondary education or the workplace into and out of public school teaching. Given that such tracking systems do not exist in California or many other states, other sources of data are used to construct indicators which can be used inferentially. Using indicators to make inferences may less satisfactory than using tracking systems. However an indicator based analysis is superior to uninformed guesswork.

A rational method for constructing indicators of the flow of individuals through institutions involves a conceptual model. One simple model includes four main components: the K-12 school system, the college/university preparation programs, the pool of re-entrants, and the pool of less than fully qualified individuals who are willing and permitted to work on an emergency basis. Increased K-12 student enrollment for example, requires hiring more teachers in order to sustain a given student-teacher ratio. The number of teachers available depends in part on enrollments in college and university preparation programs, and the capacity of colleges and universities to serve undergraduates. The pool of K-12 graduates is a source of prospective college students, and colleges/universities need enrollment to sustain degree and credential programs. Additionally, there is a continually replenished pool of former teachers, some of whom are interested in re-entering the profession. Where school districts cannot recruit and hire a sufficient number of fully qualified teachers, they turn to emergency permit hiring. While this particular model can be refined, it provides a starting point for discussion.

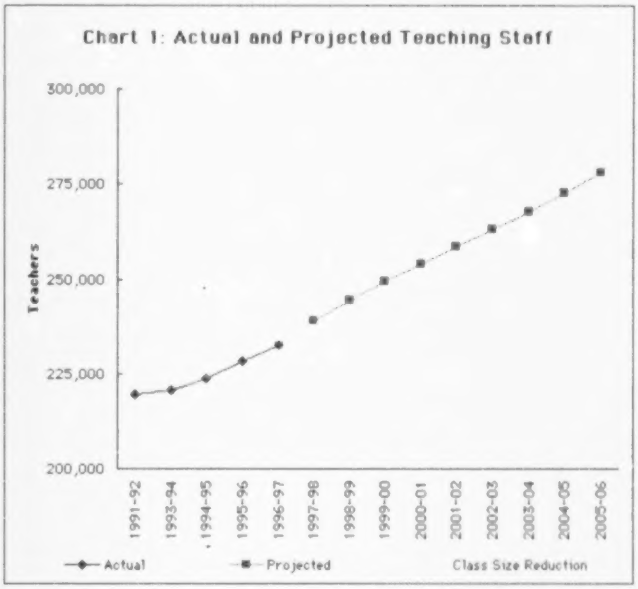
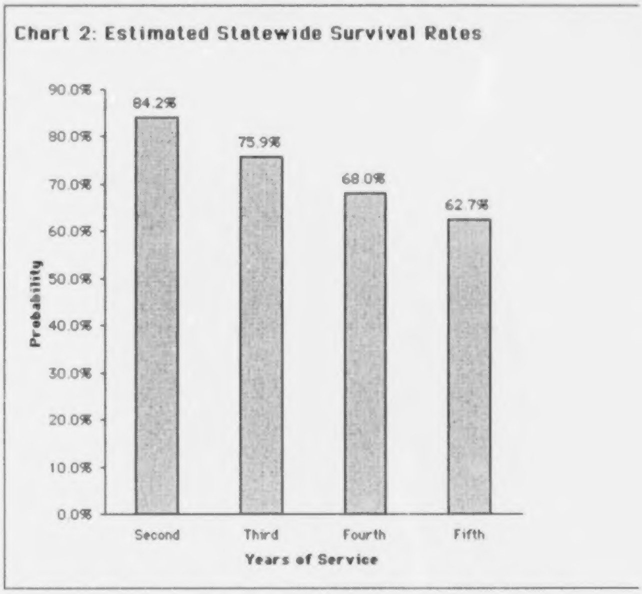


Chart 1 displays California's actual and projected graded K-12 enrollment and numbers of classroom teachers. (See Note 1.) The actual average ratio of K-12 pupils to classroom teachers from 1991-92 through 1995-96 is 23 to 1. This ratio is not a measure of class size, given that it does not take account of physical classrooms. As an overall measure, it encompasses situations with typically low student teacher ratios, such as special education, and situations with high ratios, such as some physical education classes. The projected numbers of teachers from 1996-97 onward assume continuation of the 23 to 1 student teacher ratio. During the fall of 1996 the California Legislature enacted a program, which gives incentives to school districts to reduce class size in three elementary grades. Under this program there is a limit of twenty students in a "class." An estimated 20,000 additional teachers are needed to fully implement this program, which represents about an eight percent increase in the size of the workforce. The projected number of teachers under the Class Size Reduction Program is calculated by applying an eight percent increase to the original projections. Over the next ten years, with class size reduction, the teaching workforce should increase in size by 68,000, which is about 30 percent growth.

The total number of teachers employed at any given time depends on the flows of individuals in and out of the workforce. These flows include hires of first-time or previously employed teachers, retirements, and attrition of experienced teachers. Information about retirements is available from The California State Teachers Retirement System. (See Note 2.) Between 1990 and 1995 and average of 5,150 K-12 public school teachers retired each year, typically with 28 years of service at an age of 61. The percent of staff retiring varied from a low of 1.8 percent in 1990 to a high of 2.3 percent in 1993. On average about 2 percent of teachers retired each year.

Estimates of teacher attrition can be made using data from the California Department of Education annual staff surveys of professional assignments. (See Note 3.) School districts provide the number of years of service for each teacher. A cohort of new first-time teachers is operationally defined as those who are beginning their first year of service. The size of this cohort in its second year and following years is estimated by counting the number of teachers in the district who are beginning their second year in the district and overall. The available data neither identify nor permit the longitudinal tracking of individuals, precluding a traditional application of survival analysis techniques. Even so, it is possible to develop estimates of the relevant survival statistics.



Cohorts of new first-time teachers were estimated from the surveys of professional staff for 1986-87 through 1995-96. Overall estimates of survival probabilities for this period can be made by calculating the percent of the original cohort which appears to be present in succeeding years and averaging across cohorts. By definition 100 percent of such teachers are present the first year. As displayed in Chart 2, an average of 84 percent were present at the beginning of the second year, 76 percent the third year, 68 percent the fourth year, and 63 percent the fifth year. An estimated half of all new first-time teachers remain at the beginning of a seventh year. Annual attrition of the total population of teachers is estimated at six percent. (See Note 4.) The risk or hazard that a teacher will quit in a particular year, given that he or she survived through the end of the previous year was 16 percent after the first year, 10 percent after the second, 10 percent after the third, and 9 percent after the fourth.

The estimated eight percent of teachers who leave the workforce annually due to retirement and attrition influences the annual number that must be recruited and hired. For example, in round numbers:

228,000 teachers employed in 1994-95 year
(minus) 18,000 eight percent attrition and retirement

210,000 teachers remain available for 1995-96

However, about 232,000 teachers were actually employed in 1995-96, so that 22,000 teachers were needed to fill the gap. More arithmetic indicates that about forty percent of those hired in 1995-96 were re entrants, compared to 60 percent first-time teachers.

22,000 teachers actually hired in 1995-96
(minus)13,500 actual first time teachers hired in 1995-96

8,500 re-entrants in 1995-96

Among other requirements, California teachers must possess a four-year undergraduate degree from a regionally accredited college or university, and a major in their subject area of instruction. A "fifth-year" teacher preparation program traditionally supplements the undergraduate degree. For this reason, the capacity of colleges and universities to serve undergraduates and produce undergraduate degrees influences teacher preparation. Table 1 displays annual counts of undergraduate degrees produced by the state's accredited public and independent colleges and universities along with counts of students from public high schools graduating five years earlier. (See Note 5.)

Table 1

Undergraduate Degrees and High School Graduates

Year of Degree	Number of Degrees	Graduation Year	High School Graduates
1989-90	96,270	1984-85	225,448
1990-91	99,553	1985-86	229,026
1991-92	105,446	1986-87	237,414
1992-93	108,103	1987-88	249,518
1993-94	109,850	1988-89	244,629
1994-95	107,661	1989-90	236,291
1995-96	*103,179*	1990-91	234,164
1996-97	*107,775*	1991-92	244,594
1997-98	*109,857*	1992-93	249,320
1998-99	*111,515*	1993-94	253,083
1999-00	*112,448*	1994-95	255,200
2000-01	*114,730*	1995-96	*260,378*
2001-02	*116,461*	1996-97	*264,307*
2002-03	*123,178*	1997-98	*279,552*
2003-04	*125,640*	1998-99	*285,138*

The average number of years spent pursuing an undergraduate degree is disputed, but likely depends on the type of institution (public versus private) and selectivity. Although the undergraduate degree conventionally takes four years, many students require five years or longer. Table 2 compares the number of degrees in a given year with the size of the public school graduating class five years previously. The ratio of graduates (lagged by five years) to undergraduate degrees is 0.44. Assuming that this ratio of graduates to degrees remains constant, the projections of high school graduates can be used to estimate the future supply of undergraduate degrees. Under this assumption, over the next ten years there will be an average annual net increase of 2,807 undergraduate degrees per year. The assumption of a constant ratio is probably optimistic, given a recent analysis of funding for public higher education and student aid in California showing that the state's colleges and universities probably will not receive additional resources needed to increase their capacity to serve future potential undergraduates. (Breneman, 1995)

Decreases in the number of college graduates in the late 1980s are correlated with a leveling off and actual decrease in the number of undergraduate degrees granted in recent years. The projections suggest that there will be continued decline in the number of undergraduate degrees for one more year, potentially followed by increases through 2003-04. The short term trend suggests that there will be decreasing numbers of students completing teacher preparation programs. Over the longer term, assuming that programs continue to attract and serve students at the same level as in the past, there should be increasing numbers of college and university prepared candidates. On the other hand, given the prospect of limited growth for higher education generally, it may be unrealistic to expect significantly greater numbers individuals completing teacher preparation programs.

Table 2 displays the number of newly prepared teacher candidates and the number of new teachers actually hired. (See Note 6.) New or first credentials exclude renewals, a process which California requires every five years. The number of college/university credentials includes prospective teachers who completed a college or university program. The number of emergency permits or waivers reflects people who have not met all the requirements for a credential. Out of state candidates, who are certified elsewhere, can be authorized to teach while they complete California requirements. The "Other" category includes interns and individuals who are converting types of credentials no longer in use to current ones.

Table 2

Actual New or First Time Credentials and New Hires of First Time Teachers

Year	Total	College/ University	Emergency/ Waiver	Out of State	Other	New Hires
1992-93	22,341	13,022	4,055	3,649	1,615	9,436
1993-94	22,808	13,332	5,235	3,083	1,158	12,530
1994-95	22,485	12,746	5,628	2,938	1,173	14,090
1995-96	22,767	13,432	5,408	2,700	1,226	13,535

Overall, about 58 percent of candidates were from colleges/universities, 22 percent received emergency permits or waivers, 14 percent came from out of state, and 6 percent from other sources. Emergency permits are issued only when a job offer is pending, so virtually all such recipients are employed. It is likely that teachers from out of state who seek a California credential have job offers. Candidates in the "Other" category who are converting outdated credentials may or may not be currently employed as teachers. School district interns have pending offers.

Dividing the number of new hires by the total number of credentials produces a rate which ranges from .42 to .62, averaging .55. This suggests that at most about fifty percent of all new or first time credential holders are actually hired. However, the probability of employment appears to be more remote for the college/university group, which must compete against emergency permit holders, out of state candidates and others. As stated earlier, virtually all emergency permit holders have job offers. It is not known how many out of state candidates or "other" candidates gain employment. Plausible estimates are that at least three fourths of out of state candidates and half of "others" have jobs. If true, the four year average hiring rate for college/university credential holders is .33.

The four year trends for production of college/university credentials appears to be flat. This relatively flat trend is consistent with the recent leveling off and decline in the numbers of undergraduate degrees produced. It may also be a response to the apparently rigorous competition for a limited number of jobs and the prevalence of emergency permit hiring. During the same period the number of emergency permits and waivers increased by about one third, and new hires increased by about 43 percent. The number of credentials granted to candidates from out of state has declined by about 25 percent, possibly in response to California's weak economic conditions during the first half of the decade.

Discussion

Perhaps the most basic finding is the projected increase in the size of the teaching workforce. Driven by increasing student enrollment, the number of teachers needed will grow by about thirty percent during the next ten years. Given California's compulsory school attendance laws and the virtual entitlement of the opportunity to earn an high school diploma, schools must fill these teaching slots.

Does the increased demand for teachers prophesy a shortage by the year 2005? Assuming that current rates of attrition, retirement, and hiring of re-entrants remain constant, it is difficult to make a case for shortages. At the beginning of the school year in 2005 there will be an estimated need to hire about 29,000 teachers, of whom 17,400 will be first-time credential recipients. There were over 17,000 first-time or new credentials issued in 1995-96, exclusive of emergency permits. California's improving economy in coming years will probably lure more teachers from other states, further mitigating any prospect of a shortage. On the other hand, misalignment in the supply and demand of teachers is suggested by the increasing number of emergency permits issued and the apparently low employment rates for first-time college and university prepared teachers.

An estimated 22 percent of new teachers are currently hired on an emergency basis. Reynolds (1991), the National Commission on Teaching and America's Future (1996), and Hart and Burr (1996) suggest that this practice raises concerns about protection of the public interest. One concern is whether teachers on emergency permits are as effective with students as regularly credentialed teachers. Given that teachers on emergency permit lack proof of competence in their instructional area or in teaching techniques, they may well not be as effective. A second concern is how long emergency permit teachers remain in service. A lack of preparation on top of the stress of adapting to the pressures of a school environment could result in a higher attrition rate, compared to regularly credentialed teachers. If true, then the increased numbers of emergency permits should be associated with higher levels of attrition. To the extent that higher attrition rates and lower levels of preparation impede the development of a stable and experienced workforce, they probably contribute to low student achievement, poor discipline, and other undesirable student outcomes. There are administrative costs associated with teacher attrition for example, the resources expended for recruitment, hiring, and one or more years of induction. There will probably be negative consequences for the teaching staff, including lower morale, poorer attendance, and increased disciplinary problems.

While the need for teachers is growing, a corresponding increase in the capacity of teacher training programs seems unlikely. The decreasing numbers of high school graduates and degrees produced five years previously indicate that in the near term undergraduate degrees will continue to decrease, possibly depressing the enrollments in college and university teacher preparation programs. Over the next ten years, assuming optimistically that colleges and universities will be able to serve more students, there will be an estimated 15 to 20 percent increase in the production of undergraduate degrees, corresponding to a thirty percent increase in the need for teachers. The expansion of teacher preparation program budgets at public universities would likely be difficult, particularly given present difficulties in competing for funds. In the unlikely event that more resources for teacher preparation programs are forthcoming, several years will probably be required to expand the programs. To the extent that such expansion encompasses better coordination with local school district needs, it may help to remedy emergency permit hiring.

An alternative to expanding preparation programs is to improve the survival rate of teachers on the job. Currently it appears that about half the cohort of first-time teachers survives past seven years, with greater numbers of teachers leaving the profession after the first or second year. Programs to retain more teachers for more years in the profession could reduce the need for hiring. Although such programs would be helpful statewide, they would be most useful in geographical and curricular areas that experience high rates of attrition.

An estimated one third of individuals who complete college and university preparation programs actually seek and gain regular teaching employment in public schools. The results are consistent with Tierney's (1993) finding that prospective teachers heavily weigh the closeness of jobs their current home and availability of assignment. School districts may find that the skills and abilities of the available candidates do not match the positions to be filled. On one level, the function of teacher training programs is to prepare teachers, not to guarantee employment. On the other hand, the room for improvement in the employment rate suggests a mismatch between labor market needs and the preparation programs.

Another alternative to college and university programs is the reserve pool of teachers who were previously employed in public schools and now wish to re-enter service. Over half of the teachers hired in 1995-96 had previous experience. Beaudin's (1993, 1995) findings suggest that the decision to re-enter may depend heavily on financial issues. A previously employed teacher is more likely to re-enter if the teaching job is geographically convenient and is higher paying than private sector alternatives. The hiring of re-entering teachers provides a context for interpreting the attrition of first-time teachers. While many first-timers leave after one or two years, it appears that many of them eventually return to teaching.

Conclusion

It was earlier noted that the quality of teacher preparation and the quantity of available teachers are not independent. A traditional interpretation of this statement is that higher preparation standards by limiting access threaten the supply of teachers and the staffing of schools. It would be unfortunate if the projected need for more teachers were to cause an erosion of standards for teacher preparation. This scenario leads towards lower student performance, less job satisfaction, higher teacher attrition, increased public discontent, and further erosion of standards. Easier teacher preparation programs and emergency permit hiring are expedient solutions to short term employment needs. However, such expediency may bring about greater long term problems.

Given the findings of this study a different interpretation and policy seems plausible. Teachers who are more thoroughly prepared to meet the specific needs of schools may persist longer in their jobs. If this is true, higher retention rates of qualified teachers would result in the establishment of a more stable, satisfied, and highly competent workforce, slowing the revolving employment door at school district offices, and reducing the need for emergency permit hiring. An additional, perhaps more important benefit is that better prepared teachers should be more effective in their jobs and assist more students to higher levels of attainment. A policy of higher standards and more support may be difficult to achieve in the near term. Public schools will have to be weaned from expedient employment practices, and preparation programs will need to become more rigorous and attentive to local needs. In the longer term this policy will benefit students and the teaching profession.

References

Ashton, P. (1996). Improving the preparation of teachers. Educational Researcher. Vol. 25, No. 9, pp. 21-22.

Beaudin, B. (1993). Teachers who interrupt their careers: Characteristics of those who return to the classroom. Educational Evaluation and Policy Analysis. Vol. 15 No.1, pp. 51-64.

Beaudin, B. (1995). Former teachers who return to public schools: District and teacher characteristics of teachers who return to the districts they left. Educational Evaluation and Policy Analysis. Vol. 17, No. 4, pp. 462 475.

Breneman, D. (1995). A State of Emergency? Higher Education in California. San Jose: The California Higher Education Policy Center.

Darling-Hammond, L., and Hudson, L. (1990). Precollege science and mathematics teachers: Supply, Demand, and Quality. In Review of Research in Education (Vol. 16). Washington, DC: American Educational Research Association.

Glass, G., and Smith M. (1979). Meta-analysis of research on class size and achievement. Educational Evaluation and Policy Analysis. Vol. 1, No. 1, pp. 2-16.

Greenwald, R., Hedges, L., and Laine, R. (1996). The effect of school resources on student achievement. Review of Educational Research. Vol. 66, No. 3, pp. 361-396.

Hanushek, E. (1996). A more complete picture of school resource policies. Review of Educational Research. Vol. 66, No. 3, pp. 397-409.

Hart, G. And Burr, S. (1996). A State of Emergency ... In a State of Emergency Teachers. Sacramento: California State University Institute for Education Reform.

Kirst, M., Hayward, G., and Koppich, J. (1995). Conditions of Education in California 1994-95. Policy Paper No. PP95-4-1. Berkeley, California: Policy Analysis for California Education (PACE), University of California.

Mosteller, F., Light, R., and Sachs, J. (1996). Sustained inquiry in education: Lessons from skill grouping and class size. Harvard Educational Review, Vol. 66, No. 4, Winter 1996, pp. 797-842.

National Center for Educational Statistics. (1996a). The Condition of Education 1996. Washington DC: Author.

National Center for Educational Statistics. (1996b). Out-of-Field Teaching and Educational Equality. Washington DC: Author.

National Center for Educational Statistics. (1996c). Projections of Education Statistics to 2006. Washington DC: Author.

National Commission on Teaching and America's Future (1996). What Matters Most: Teaching for America's Future. New York, Teachers College, Columbia University.

Reynolds, A. (1991). What is competent beginning teaching? A review of the literature. Review of Educational Research. Vol 62, No. 1, pp 1-35.

Tierney, D. (1993). A study of the employment patterns of recent graduates of California teacher education programs and the employment decisions of a selected sample of California school districts. Unpublished study. Sacramento: California Commission on Teacher Credentialing.

Willet. J., and Singer, J. (1991). From whether to when: New methods for studying student dropout and teacher attrition. Review of Educational Research. Vol 61, No. 4, pp. 407-450.

Notes

1. The actual graded K-12 enrollment and numbers of teachers shown in Table 3 are from the California Basic Educational Data System (CBEDS) of the California Department of Education. The enrollment figures are published annually in a document entitled "California Public Schools Enrollment." The numbers of teachers are published annually in a document entitled "Count of Certificated and Classified Staff in California Public School Districts." The teacher counts reflect certificated staff with classroom assignments, and exclude administrators and pupil services staff.

Table 3

Actual and Projected Enrollment and Teaching Staff

School Year	Graded K-12 Enrollment	Teachers	Class Size Reduction Program
1991-92	5,001,670	219,353	
1992-93	5,089,808	220,871	
1993-94	5,166,261	223,932	
1994-95	5,242,078	228,204	
1995-96	5,367,926	232,488	
1996-97	*5,495,075*	*238,951*	*259,000*
1997-98	*5,623,422*	*244,532*	*266,000*
1998-99	*5,737,874*	*249,509*	*269,000*
1999-00	*5,841,535*	*254,017*	*274,000*
2000-01	*5,945,067*	*258,519*	*279,000*
2001-02	*6,052,242*	*263,179*	*284,000*
2002-03	*6,160,231*	*267,875*	*289,000*
2003-04	*6,271,881*	*272,730*	*295,000*
2004-05	*6,392,367*	*277,969*	*300,000*

The projections of graded public school enrollment are published by the California Department of Finance Demographic Research Unit in a document entitled "K-12 Graded Public School Enrollment by Ethnicity, History, and Projection - 1995 Series." The projections are based on a grade-progression ratio (or cohort survival) projection method and the most recent ten years of historical enrollment data from CBEDS.

2. Counts of new retirants from K-12 districts from 1990 through 1995 were obtained by special request in February, 1996 from the California State Teachers Retirement System (CSTRS). The counts are cumulated over fiscal years ending June 30. Unlike the counts regularly published in

the CSTRS annual report which include community college district staff, these data reflect certificated staff in K-12 districts only. Teachers make up 92 percent of all employed certificated staff according to published California Department of Education statistics. Therefore the number of retired teachers is estimated as 92 percent of all K-12 retired certificated staff counted by CSTRS.

3. Counts of teachers were obtained from a direct tabulation of the results of the annual Professional Assignment Information Form (PAIF), an annual survey conducted as a part of CBEDS. The information requested on the PAIF is required of each certificated staff, and includes demographics, assignments, and position/credentials. A "first-time" teacher is operationally defined as someone whose primary assignment is teaching and who reports being both new to the teaching profession and new to the district.

The CBEDS PAIF file does not include individual identifiers so that teachers cannot be tracked from year to year. It is possible to estimate cohort attrition by counting the number of teachers who report being their second year both at the district and in the profession, in their third year, fourth year, and so on. Unfortunately, this method does not account for those teachers who transfer from one school district to another, re entrants, or those who accept administrative jobs. The method of estimating attrition therefore undercounts the number of teachers who remain in the profession. On the other hand, when a teacher accepts a promotion or leaves for another district, a school district must recruit, select, and induct another teacher in order to fill the vacancy. Cross district mobility within the teaching profession has costs which are similar to those caused by attrition.

4. Table 4 contains actual counts of the number of teachers who reported that they were new to service and new to a district, who reported two years of service in a district and overall, and so on, up through five years. The counts are arranged so that the size of an hypothetical cohort can be followed across a row. The Professional Assignment Information Survey was not conducted in 1992, so data from that year are not available. The consequence is that survival and hazard rates for the 1992 cohort cannot be estimated.

Table 4

Actual Counts of Teachers by Years of Service

Cohort Year	First Year	Second Year	Third Year	Fourth Year	Fifth Year
1987	9,025	7,317	6,450	5,729	5,235
1988	8,301	7,242	6,433	5,865	n/a
1989	9,446	8,151	7,212	n/a	6,208
1990	11,523	10,076	n/a	8,002	7,471
1991	12,153	n/a	8,930	8,275	7,505
1992	n/a	7,658	7,235	6,482	5,792
1993	9,436	8,459	7,331	6,444	
1994	12,530	10,439	9,916		
1995	14,090	10,504			
1996	13,528				

An exponential regression procedure was used to estimate the number of years the average teacher remains in the classroom. The independent variable was the number of years survived, ranging from 2 to 5. The dependent variable was the probability of survival associated with that particular number of years. Each cohort furnished up to four pairs of such numbers, depending on the starting year, which ranged from 1986-87 through 1995-96. The Excel logest function estimated the parameters for the model Y = B*M^X, with B = 1.026, and M = 0.904. The regression model estimates retention at 83.9 percent at the beginning of the second year, 75.9 at the beginning of the third, 68.6 for the fourth, 62.0 for the fifth, 50.7 for the fifth, and 45.9 for the sixth. It appears that about half of the cohort remains at the beginning of the seventh year.

Similar techniques can be used to estimate the annual rate of teacher attrition. Empirical hazard probabilities (the conditional probability that a teacher will leave, given that he or she survived through the end of the previous year) are calculated and used to fit an exponential regression model similar to the one used above to estimate survival rates. The parameter estimates for this model were B = 0.145, and M = 0.892. The fitted regression models are used to estimate survival and hazard probabilities for years 2 through 30. A weighted average of the hazard probabilities is calculated, using the survival probabilities as weights. The survival probability is interpreted as an estimate of the percentage contribution of an historical cohort to the current population of teachers. The associated hazard probability is the likelihood of attrition from that cohort. The weighted average was 6.0 percent, estimating the annual rate of teacher attrition. One caution is that this estimate relies on extrapolation of statistics beyond the range of available data. On the other hand, the result is identical to the annual attrition rate cited by Willett and Singer (1991, p. 410) for the United States as a whole and in various individual states.

5. The numbers of undergraduate degrees from 1989-90 through 1994-95 are taken from a report, "The Performance of California Higher Education," produced annually by the California Postsecondary Education Commission. The counts include degrees produced by the University of California, the California State University, and by accredited California independent colleges and universities.

The actual counts of high school graduates are taken from CBEDS K-12 enrollment reports, as described in Note 1. The projections of high school graduates are published by the California Department of Finance Demographic Research Unit in a document entitled "K-12 Public High School Graduates by Ethnicity, History, and Projection - 1995 Series." The projections are based on a grade-progression ratio (or cohort survival) projection method and the most recent ten years of historical enrollment data from CBEDS.

An unknown number of students who enroll at California colleges and universities and earn undergraduate degrees come from private schools in California, or from schools out of state. Equally unknown is the number of people who complete an undergraduate degree in California and leave the state, or who complete a degree elsewhere and move to California. Even so, it appears that the bulk of teachers employed in California also received their undergraduate education in-state. Table 3, which includes counts of out-of-state applicants appears to support this assumption.

The ratio of graduates to degrees, while providing a basis for making projections, simplifies a complex situation. Production of undergraduate degrees depends not only the number of potential degree seekers, but also on the level of resources that colleges and universities have to serve students, and on the goals of high school graduates. The ratio appears to be quite stable over the years covered in Table 3, and the correlation between graduates and degrees is quite stable.

6. The counts of first time and new credentials were obtained directly from the Credential Automation System, a database which is maintained by the California Commission on Teacher Credentialing. Although records are kept of credential and permit issuances, no records are kept of actual public school employment. The number of new or first-time credentials is an approximate count of the number of people receiving credentials. A few individuals may receive more than one new or first-time credential. It is possible that a person could receive a credential to work in one area, but then receive an emergency permit to teach in another subject. Although most people with emergency permits have pending job offers, it is also possible that some may not actually report for work.

About the Author

Mark Eric Fetler, Ph.D.

California Commission on Teacher Credentialing
1812 Ninth Street
Sacramento, CA 95814

(916) 445-3223
mfetler@ctc.ca.gov

Employment Experience

- Consultant. California Commission on Teacher Credentialing. 1995 - present. Plan, organize and conduct research on teacher credential examinations, including the California Basic Education Skills Test. Contract management for the Multiple Subjects Assessment for Teachers. Develop and track indicators of teacher supply and demand.
- Director. Planning, Effectiveness and Accountability Unit, Chancellor's Office, California Community Colleges. 1990 - 1995. Manage accountability task force. Develop federal and state accountability programs.
- Administrator. Educational Planning and Information Center, California Department of Education. 1984 - 1990. Manage and develop K-12 accountability policies and programs.
- Consultant. California Assessment Program, California Department of Education. 1980 - 1984. Conduct research and assessment projects.
- Evaluation Specialist. Northwest Regional Educational Laboratory, Portland, Oregon. 1979 - 1980. Deliver evaluation technical assistance and training to state education agencies.
- Senior Staff Associate. Western Interstate Commission for Higher Education, Boulder, Colorado, 1978 - 1979. Evaluate research instruments.

Education

B.A., Pyschology, Colorado College, 1972

Ph.D., Psychology, University of Colorado, 1978

Study abroad in Germany at the Universities of Göttingen and Bielefeld

Publications

Authored research in peer-reviewed journals, e.g., *American Educational Research Journal, Educational Evaluation and Policy Analysis, Applied Measurement in Education, Sex Roles*, and the *Journal of Communication.*

Community Service

Volunteer for Elk Grove Unified School District, Board member of the Elk Grove Community Planning Advisory Council, fund raiser for the Strauss Festival of Elk Grove, and member of the Elk Grove Rotary Club.

Copyright 1997 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://olam.ed.asu.edu/epaa

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey@olam.ed.asu.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Marshall University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Rocky Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
Arizona State University

Dennis Sayers
University of California at Davis

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina--Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher Education

Mary P. McKeown
Arizona Board of Regents

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois--UC

Robert T. Stout
Arizona State University

archives | abstracts | editors | board | submit | comment | subscribe | search

• enter the archives • browse the abstracts • the editors • the edit board
• submit article • submit commentary • search • subscribe
volume: • 1 • 2 • 3 • 4 • 5 • 6 • 7

This article has been retrieved 6478 times since January 15, 1997.

Education Policy Analysis Archives

Volume 5 Number 3 **January 15, 1997** **ISSN 1068-2341**

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass Glass@ASU.EDU.

College of Education

Arizona State University,Tempe AZ 85287-2411

Copyright 1997, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

Testing Writing on Computers: An Experiment Comparing Student Performance on Tests Conducted via Computer and via Paper-and-Pencil

Michael Russell
Boston College

Walt Haney
Boston College

Abstract

Computer use has grown rapidly during the past decade. Within the educational community, interest in authentic assessment has also increased. To enhance the authenticity of tests of writing, as well as of other knowledge and skills, some assessments require students to respond in written form via paper-and-pencil. However, as increasing numbers of students grow accustomed to writing on computers, these assessments may yield underestimates of students' writing abilities. This article presents the findings of a small study examining the effect that mode of administration -- computer versus paper-and-pencil -- has on middle school students' performance on multiple-choice and written test questions. Findings show that, though multiple-choice test results do not differ much by mode of administration, for students accustomed to writing on computer, responses written on computer are substantially higher than those written by hand (effect size of 0.9 and relative success rates of 67% versus 30%). Implications are discussed in terms of both future research and test validity.

Introduction

Two of the most prominent movements in education over the last decade or so are the introduction of computers into schools and the increasing use of "authentic assessments." A key assumption of the authentic assessment movement is that instead of simply relying on multiple choice tests, assessments should be based on the responses students generate for open-ended "real world" tasks. "Efforts at both the national and state levels are now directed at greater use of performance assessment, constructed response questions and portfolios based on actual student work" (Barton & Coley, 1994, p. 3). At the state level, the most commonly employed kind of non-multiple-choice test has been the writing test (Barton & Coley, 1994, p. 31) in which students write their answers long-hand. At the same time, many test developers have explored the use of computer administered tests, but this form of testing has been limited almost exclusively to multiple-choice tests. Relatively little attention has been paid to the use of computers to administer tests which require students to generate responses to open-ended items.

The consequences of the incongruities in these developments may be substantial. As the use of computers in schools and homes increases and students do more of their writing with word processors, at least two problems arise. First, performance tests which require students to produce responses long-hand via paper-and-pencil (which happens not just with large scale tests of writing, but also for assessments of other skills as evidenced through writing) may violate one of the key assumptions of the authentic assessment movement. For people who do most of their writing via computer, writing long-hand via paper-and-pencil is an artificial rather then real world task. Second, and more importantly, paper-and-pencil tests which require answers to be written long-hand to assess students' abilities (in writing or in other subjects) may yield underestimates of the actual abilities of students who are accustomed to writing via computer.

In this article, we present the results of a small study on the effect of computer administration on student performance on writing or essay tests. Specifically, we discuss the background, design and results of the study reported here. However, before focusing on the study itself, we present a brief summary of recent developments in computerized testing and authentic assessment.

In 1968, Bert Green, Jr., predicted "the inevitable computer conquest of testing" (Green, 1970, p. 194). Since then, other observers have envisioned a future in which "calibrated measures embedded in a curriculum . . . continuously and unobtrusively estimate dynamic changes in student proficiency" (Bunderson, Inouye & Olsen, 1989, p. 387). Such visions of computerized testing, however, are far from present reality. Instead, most recent research on computerized testing has focused on computerized adaptive testing, typically employing multiple-choice tests. Perhaps the most widely publicized application of this form of testing occurred in 1993 when the Graduate Record Examination (GRE) was administered nationally in both paper/pencil and computerized adaptive forms.

Naturally, the introduction of computer administered tests has raised concern about the equivalence of scores yielded via computer- versus paper-and-pencil-administered test versions. Although exceptions have been found, Bunderson, Inouye & Olsen (1989) summarize the general pattern of findings from several studies which examined the equivalence of scores acquired through computer or paper- and-pencil test forms as follows: "In general it was found more frequently that the mean scores were not equivalent than that they were equivalent; that is the scores on tests administered on paper were more often higher than on computer-administered tests." However, the authors also state that "[t]he score differences were generally quite small and of little practical significance" (p. 378). More recently, Mead & Drasgow (1993) reported on a meta-analysis of 29 previous studies of the equivalence of computerized and paper-and-pencil cognitive ability tests (involving 159 correlations between computerized and paper-and-pencil test results). Though they found that computerized tests were slightly harder than paper-and-pencil tests (with an overall cross-mode effect size of -.04), they concluded that their results "provide strong support for the conclusion that there is no medium effect for carefully constructed power tests. Moreover, no effect was found for adaptivity. On the other hand, a substantial medium effect was found for speeded tests" (Mead & Drasgow, 1993, p. 457).

Yet, as previously noted, standardized multiple-choice tests, which have been the object of comparison in previous research on computerized versus paper-and-pencil testing, have been criticized by proponents of authentic assessment. Among the characteristics which lend authenticity to an assessment instrument, Darling-Hammond, Ancess & Falk (1995) argue that the tasks be "connected to students' lives and to their learning experiences..." and that they provide insight into "students' abilities to perform 'real world' tasks" (p.4-5). Unlike standardized tests, which may be viewed as external instruments that measure a fraction of what students have learned, authentic assessments are intended to be closely linked with daily classroom activity so that they seamlessly "support and transform the process of teaching and learning" (Darling-Hammond, Ancess & Falk, 1995, p. 4; Cohen, 1990).

In response to this move towards authentic assessment, many developers of nationally administered standardized tests have attempted to embellish their instruments by including open-ended items for which students have to write their answers. These changes, however, have occurred during a period when both the real-world and the school-world have experienced a rapid increase in the use of computers.

The National Center for Education Statistics report that the percentage of students in grades 1 to 8 using computers in school has increased from 31.5 in 1984, to 52.3 in 1989 and to 68.9 in 1993 (Snyder & Hoffman, 1990; 1994). In the workplace, the percentage of employees using computers has risen from 36.0 in 1989 to 45.8 in 1993. During this period, writing has been the predominant task adult workers perform on a computer (Snyder & Hoffman, 1993; 1995). Given these trends, tests which require students to answer open-ended items via paper-and-pencil may decrease the test's "authenticity" in two ways: 1. Assessments are not aligned with students' learning experiences; and 2. Assessments are not representative of 'real-world' tasks. As the remainder of this paper suggests, these shortcomings may be leading to underestimates of students' writing abilities.

Background to this Study

In 1993, the Advanced Learning Laboratory School (ALL School, (http://nis.accel.worc.k12.ma.us) of Worcester, Massachusetts decided to adopt the Co-NECT school design (or Cooperative Networked Educational Community for Tomorrow, http://co-nect.bbn.com). Developed by BBN, Inc., a Boston- based communications technology firm, Co-NECT is one of nine models for innovative schooling funded by the New American Schools Development Corporation. Working with BBN, the ALL School restructured many aspects of its educational environment. Among other reforms, the traditional middle school grade structure (that is, separately organized grade 6, 7 and 8 classes) was replaced with blocks which combined into a single cluster students who otherwise would be divided into grades 6, 7 and 8. In place of traditional subject-based classes (such as English Class, Math Class, Social Studies, etc.), all subjects were integrated and taught through project-based activities. To support this cooperative learning structure, several networked computers were placed in each classroom, allowing students to perform research via the Internet and CD-ROM titles, to write reports, papers and journals, and to create computer based presentations using several software applications.

To help evaluate the effects the restructuring at the ALL School has on its students as a whole, the Center for the Study of Testing, Evaluation and Educational Policy (CSTEEP) at Boston College helped teachers gather baseline data in the fall of 1993 with plans to perform follow-up assessments in the spring of 1994 and each spring thereafter. To acquire a broad picture of students' strengths and weaknesses, the forms of tests included in the baseline assessment ranged from multiple choice tests to short and long answer open-ended assessments to hands-on performance assessments covering a wide range of reading, writing, science and math skills. To acquire insight into how cooperative projects affected the development of group skills, some of the performance assessments required students to work together to solve a problem and/or answer specific questions. Finally, to evaluate how the Co-NECT Model, as implemented in the ALL School, affected students' feelings about their school, a student survey was administered. Assessments and surveys were administered to representative samples of the whole school's student population.

In the spring of 1994, the same set of assessments was re-administered to different representative samples of students. While a full discussion of the results is beyond the scope of this paper, many of the resulting patterns of change were as expected. For example, performance items which required students to work cooperatively generally showed more improvement than items which required students to work independently. On items that required students to work independently, improvement was generally stronger on open-ended items than on multiple-choice items. But there was one notable exception: open-ended assessments of writing skills suggested that writing skills had declined.

Although teachers believed that the Co-NECT Model enhanced opportunities for students to practice writing, performance on both short answer and long answer writing items showed substantial decreases. For example, on a short answer item which asked students to write a recipe for peace, the percentage of students who responded satisfactorily decreased from 69% to 51%. On a long answer item which asked students to imagine a superhero, describe his/her powers, and write a passage in which the superhero uses his/her powers, the percentage of satisfactory responses dropped from 71% to 41%. On another long answer item that asked students to write a story about a special activity done with their friends or family, student performance dropped from 56% to 43%. And on a performance writing item which first asked students to discuss what they saw in a mural with their peers and then asked them to write a passage independently that described an element in the mural and explain why they selected it, the percentage of satisfactory responses decreased from 62% to 47%. These declines were all statistically significant, and more importantly were substantively troubling.

Since writing was a skill the school had selected as a focus area for the 1993-94 school year, teachers were surprised and troubled by the apparent decrease in writing performance. During a feedback session on results in June 1994, teachers and administrators discussed at length the various writing activities they had undertaken over the past year. Based on these conversations, it was evident that students were regularly presented with opportunities to practice their writing skills. But a consistent comment was that teachers in the ALL School were increasingly encouraging students to use computers and word processing tools in their writing. As several computers were present in all classrooms, as well as in the library, teachers believed that students had become accustomed to writing on the computer. When one teacher suggested that the decrease in writing scores might be due to the fact that all writing items in spring 1994 were administered on paper and required students to write their responses by hand, the theory was quickly supported by many teachers. With a follow-up assessment scheduled to occur a year later, several teachers asked if it would was possible for students to perform the writing items on a computer.

After careful consideration, it was decided that a sub- sample of students in spring 1995 would perform a computer- administered version of the performance writing item and items from the National Assessment of Educational Progress (NAEP) (items were mostly multiple-choice with a few short answer items included). But, to preserve comparisons with results from 1993-94, the majority of the student population would perform these assessments as they had in that year -- via the traditional pencil-and-paper medium. Hence, we undertook an experiment to compare the effect that the medium of administration (computer versus paper-and-pencil) has on student performance on multiple-choice, short-answer and extended writing test items.

Study Design and Test Instruments

To study the effect the medium of administration has on student performance, that is taking assessments on computer versus by hand on paper, two groups of students were randomly selected from the ALL School Advanced Cluster (grades 6, 7 and 8). For the experimental group, which performed two of three kinds of assessments on computer, 50 students were selected. The control group, which performed all tests via pencil-and-paper, was composed of the 70 students required for the time-trend study described above. The three kinds of assessments performed by both groups were:

1. An open-ended (OE) assessment comprising 14 items, which included two writing items, five science items, five math items and two reading items.
2. A test comprised of NAEP items which was divided into three sections and included 15 language arts items, 23 science items and 18 math items. The majority of NAEP items were multiple-choice. However, 2 language arts items, 3 science items and 1 math item were open-ended and required students to write a brief response to each item's prompt.
3. A performance writing assessment which required an extended written response.

Both groups performed the open-ended (OE) assessment in exactly the same manner, by hand via paper-and pencil. The experimental group performed the NAEP and writing assessment on computer, whereas the control group performed both in the traditional manner, by hand on paper.

The performance writing assessment consisted of a picture of a mural and two questions. Students formed small groups of 2 or 3 to discuss the mural. After 5 to 10 minutes, students returned to their seats and responded to one of two prompts:

1. Now, it is your turn to pick one thing you found in the mural. Pick one thing that is familiar to you, that you can recognize from your daily life or that is part of your culture. Describe it in detail and explain why you chose it.
2. Artists usually try to tell us something through their paintings and drawings. They may want to tell us about their lives, their culture or their feelings about what is happening in the neighborhood, community or world. What do you think the artists who made this mural want to tell us? What is this mural's message?

Due to absences, the actual number of students who participated in this study was as follows:

Experimental (Computer) Group: 46
Control (Paper-and-Pencil) Group: 68

It should be noted that the study described in this paper was performed as part of a larger longitudinal study which relied heavily on matrix sampling. For this reason, not all of the students in the control group performed all three tests. However, all students included in the analyses reported here performed at least two tests, one of which was the open-ended assessment. Table 1 shows the actual number of students in each group that performed each test.

Table 1
Number of Students Performing Each Test

Test	Experimental	Control	Total
Open-ended	46	68	114
NAEP	44	42	86
Perf. Writing	40	46	86

To be clear, we emphasize that the treatment, in terms of which the experimental and control groups differed, had nothing to do with educational experience of the two groups. The groups were receiving similar -- albeit quite unusual in comparison to most middle schools -- educational experiences in the ALL school. The treatment, in terms of which the two groups differed, was simply that the experimental group took the NAEP and performance writing tests on computer, whereas the control group took these tests in the traditional manner, by hand with paper-and-pencil.

Converting Paper Tests to Computer

Before the tests could be administered on computer, the paper versions were converted to a computerized format. Several studies suggest that slight changes in the appearance of an item can affect performance on that item. Something as simple as changing the font in which a question is written, the order items are presented, or the order of response options can affect performance on that item (Beaton & Zwick, 1990; Cizek, 1991). Other studies have shown that people become more fatigued when reading text on a computer screen than when they read the same text on paper (Mourant, Lakshmanan & Chantadisai, 1981). One study (Haas & Hayes, 1986) found that when dealing with passages that covered more than one page, computer administration yielded lower scores than paper-and-pencil administration, apparently due to the difficulty of reading extended text on screen. Clearly, by converting items from paper to computer, the appearance of items is altered.

To minimize such effects, each page of the paper version of the NAEP items and the performance writing item was replicated on the computer screen as precisely as possible. To that end, the layout of text and graphics on the computer version matched the paper version, including the number of items on a page, the arrangement of response options, and the positioning of footers, headers and directions. Despite these efforts, not every screen matched every page. Since the computer screen contained less vertical space, it was not always possible to fit the same number of questions on the screen as appeared on the page. In addition, to allow the test taker to move between screens (e.g., to go on to the next screen, back to a previous screen, or to flip to a passage or image to which an item referred), each screen of the computer versions contained navigation buttons along its bottom edge. Finally, to decrease the impact of screen fatigue, a larger font was used on the computer version than on the paper version.

To create a computerized version of the NAEP and performance writing tests, the following steps were taken:

1. An appropriate authoring tool was selected. To fully integrate the several graphics used in the multiple-choice items and the full-color photograph of a mural used in the performance writing item, as well as to track students' responses, Macromedia Director was used.
2. All graphics and the photograph of the mural were scanned. Adobe Photoshop was used to retouch the images.
3. A data file was created to store student input, including name, ID number, school name, birth date, gender, date of administration and responses to each item.
4. A prototype of each test was created, integrating the graphics, text and database into a seamless application. As described earlier, navigational buttons were placed along the lower edge of the screen. In addition, a "cover" page was created in which students entered

biographical information.

5. The prototype was tested on several adults and students to assure that all navigational buttons functioned properly, that data was stored accurately, and that items and graphics were easy to read.
6. Finally, the prototype was revised as needed and the final versions of the computer tests were installed on twenty-four computers in the ALL School.

As described above, the addition of navigational buttons along the lower edge of the computer screen was the most noticeable difference between the computer and paper versions of the tests. To allow students to review their work and make changes as desired, a "Next Page" and "Previous Page" button appeared on all pages (or screens) of the computer tests (except the first and last page). To allow students to review their work, student responses were not recorded until the student reached the last page of the assessment and clicked a button labeled "I'm Finished." When the "I'm Finished" button was clicked, the student's biographical information and responses to each item were recorded in a data file before the program terminated. For all multiple-choice items, students clicked the option they felt best answered the question posed. For both short- and long-answer questions, examinees used a keyboard to type their answers into text boxes which appeared on their screen. Though they could edit using the keyboard and mouse, examinees did not have access to word processing tools such as spell-checking.

Scoring

A combination of multiple choice and open-ended items were performed by both groups of students. Multiple-choice NAEP items were scored as either correct or incorrect based upon the answer key accompanying the NAEP items. To prevent rater bias based on the mode of response, all short-answer NAEP responses were entered verbatim into the computer. Responses of students who had taken the NAEP questions on computer and via paper-and-pencil were then randomly intermixed. Applying the rating rubrics designed by NAEP, two raters independently scored each set of six short answer items for each student. As part of an overall strategy to summarize results on all items in terms of percent correct, the initial ratings (which ranged from 1 - 5) were converted to a dichotomous value: 1 or 0; to denote whether student responses were adequate or inadequate. The two raters' converted scores were then compared. Where discrepancies occurred, the raters re-evaluated responses and reached consensus on a score.

To score the performance writing item, all hand written responses were entered verbatim into the computer -- again so as to prevent raters from knowing which responses were originally written by hand. The hand-written and computer- written responses were randomly intermixed. Three independent raters then scored each written response, using the following four-point scoring rubric:

1. Too brief to evaluate: Student did not make an attempt; indicates that student either did not know how to begin, or could not approach the problem in an appropriate manner.
2. Inadequate Response: Student made an attempt but the response was incorrect, reflected a misconception and/or was poorly communicated.
3. Adequate Response: Response is correct and communicated satisfactorily, but lacks clarity, elaboration and supporting evidence.
4. Excellent Response: Response is correct, communicated clearly and contains evidence which supports his/her response.

Initial analyses of the three raters' ratings showed that there was only a modest level of inter-rater reliability among the three (inter-rater correlations ranged from 0.44 to 0.62, across the total of 89 performance writing responses). Although these correlations were lower than expected, research on the assessment of writing has shown that rating of writing samples, even among trained raters, tends to be only modestly reliable (Dunbar, Koretz, & Hoover, 1991). Indeed, that is why we planned to have more than one rater evaluate each student response to the performance writing task. Hence for the purpose of the study reported here we created composite performance rating scores by averaging the three ratings of each student's response (which we call PWAvg).

Since the open-ended assessment was performed by paper- and-pencil by all students, student responses were not entered into the computer. A single rater, who did not know which students had performed other assessments on the computer, scored all responses using a 4 point scale. Although each of the 14 items had its own specific scoring criteria, the general meaning of each score was the same across all 14 open-ended items, as well as the performance writing item. The raw scores were then collapsed into a 0, 1 scale, with original scores of 1 or 2 representing a 0, or inadequate response, and original scores of 3 or 4 representing a 1, or adequate response. For the purpose of the study reported here, total open-ended response scores were calculated by summing across all 14 OE items.

Results

In presenting results from this study, we discuss: 1) assessment results overall; 2) comparative results from the two groups that took assessments via computer or via paper-and-pencil; 3) results of regression analyses; and 4) separate analyses of performance on the short-answer and multiple-choice NAEP items.

We present descriptive data summaries before results of statistical tests. Regarding the latter, we note that this experiment involved multiple comparisons of results based on just two random samples of students. While the literature on how to adjust alpha levels to account for multiple comparisons (e.g. Hancock & Klockars, 1996) is too extensive to review here, let us simply summarize how we dealt with this issue. We planned to compare results for the experimental and control groups on five different measures: OE, performance writing, and three NAEP subtests, in science, math, and language arts. The Dunn approach to multiple comparisons tells us that the α for *c* multiple comparisons, α_{pc}, is related to simple α for a single comparison, as follows:

$$\alpha_{pc} = 1 - (1-\alpha)^{1/c}$$

Hence for five comparisons, the adjusted value of a simple 0.05 alpha level becomes 0.0102. Analogously, a simple alpha level of 0.01 for a single comparison becomes 0.0020 for five planned comparisons. We use these alpha levels in discussing the statistical significance of comparisons between experimental and control group results. In discussion, we address not just the statistical significance, but also the substantive significance of our findings.

Overall Results

The actual raw data on which all analyses are based is being made available to the reader. From this point, the data files can be accessed in ASCII or EXCEL Spreadsheet (binary) form.

Table 2 presents a summary of overall results, that is, combined results for all students who took any of the three assessments in Spring 1995.

Table 2
Summary Statistics for All Assessments

	Scale Range	n	Mean	SD
OE	0-14	114	7.87	2.96
NAEP Lang Arts	0-15	86	9.84	3.79
NAEP Science	0-23	86	9.70	4.37
NAEP Math	0-18	86	6.21	3.39
Perf Writing Avg	1-4	86	2.53	0.62

These data indicate that the assessments were relatively challenging for the students who performed them. Mean scores were in the range of 56-66% correct for the OE and NAEP Language Arts tests, but considerably below 50% correct for the NAEP science and NAEP math subtests. In this regard, it should be noted that all of these assessments were originally designed to be administered to eighth graders, but in the study reported here they were administered to 6th, 7th and 8th grade level students who in the ALL school are intermixed in the same clusters.

Table 3 presents Spearman rank order intercorrelations of all assessments, again across both groups. The OE results correlated only slightly higher with the PWAvg results, possibly reflecting the fact that both of these assessments were open-ended requiring students to produce rather than select an answer. The three NAEP item subtests showed moderate intercorrelations (0.56-0.62) which might be expected for multiple-choice tests in the different subject areas (despite the fact that none of the NAEP subtests contained as many as two dozen items). The PWAvg results showed modest correlations with the NAEP subtests. Of the three NAEP sub-tests, the PWAvg was most strongly correlated with the Science sub-test. Although the NAEP science results were based largely on multiple choice items, of the three NAEP subtests, the Science section contained the largest number of short answer items (3 out of 23 items). The NAEP subtest that correlated least with the PWAvg scores (0.37) was the NAEP Math subtest, which contained only one open-ended item.

Table 3
Intercorrelations of Assessment Results

	OE	NAEP Lang Arts	NAEP Science	NAEP Math	Perf. Writing
OE	1.00				
NAEP Lang Arts	0.46	1.00			
NAEP Science	0.44	0.62	1.00		
NAEP Math	0.40	0.56	0.57	1.00	
Perf Writing	0.48	0.49	0.54	0.37	1.00

p <.01 for all intercorrelations

Computer versus Paper-and-Pencil Results

Table 4 presents results separately for the experimental and control groups, namely the group which took NAEP and performance writing assessments on paper and the one that took them on computer. The table also shows results of t-tests (for independent samples, assuming equal variances for the two samples and hence using a pooled variance estimate). As an aid to interpretation, the table also shows the effect of computer administration in terms of Glass's delta effect size, that is the mean of the experimental group minus the mean of the control group divided by the standard deviation of the control group. While other methods for calculating effect size have been proposed (Rosenthal, 1994, p. 237), note that results would not differ dramatically if a pooled standard deviation were used instead of the control group standard deviation.

Results indicate that, after adjusting for the planned multiple comparisons, the effect of computer administration was significant only for the PWAvg. The effect size of computer administration on the performance writing task was 0.94.

The four tests which did not show a statistically significant difference between the two groups were the OE test and the NAEP Language Arts, Science, and Math tests. The absence of a statistically significant difference on the OE test was, of course, expected since the OE test was the one test that was administered in the same form (paper-and- pencil) to the two groups. Similarly, since the NAEP tests were primarily composed of multiple-choice items, which previous research suggests are affected minimally by the mode of administration, differences between the two groups on the NAEP tests were not expected. Note however that the size of the difference in OE scores between the two groups was surprisingly large, given that the two groups had been randomly selected. The absence of four students randomly selected for the experimental group who did not take any tests may partially explain this difference. Nevertheless to explore the possibility that group differences may partially account for apparent mode of administration effects (and also, of course, to estimate effects more precisely), regression analyses were conducted.

Table 4
Summary Results by Group

	Control			*Experimental*					
	n	Mean	SD	n	Mean	SD	Effect Size (df)	t	Sig
OE	68	7.62	3.14	46	8.24	2.66	0.20 (112)	1.10	0.27
Lang Arts	42	9.24	3.96	44	3.58	0.30	0.30 (84)	1.44	0.15
Science	42	8.67	4.17	44	10.68	4.39	0.48 (84)	2.18	0.03
Math	42	6.00	3.30	44	6.41	3.51	0.12 (84)	0.56	0.58
Perf Writ.	46	2.30	0.55	40	2.81	0.59	0.94 (84)	4.16	<.0001**

** statistically significant at the 0.01 level after taking multiple comparisons into account

Regression Analyses

As a further step in examining the effects of mode of administration, regression analyses were conducted using the OE scores as a covariate and then introducing a dummy variable (0= paper/pencil group; 1= computer administration group) to estimate the effects of mode of administration on the NAEP Language Arts, Science and Math subtests and on the PWAvg scores. Results of these regression analyses are shown in Table 5.

Table 5
Results of Regression Analyses

Dependent Variable	Coeff	SE	t-ratio	Sig
NAEP Lang Arts				
Constant	5.03	1.09	4.60	<.0001**
OE	0.57	0.13	4.40	<.0001**
Group*	0.66	0.75	0.89	0.38
NAEP Science				
Constant	3.72	1.23	3.02	.0033
OE	0.67	0.15	4.59	<.0001**
Group*	1.42	0.84	1.69	0.09
NAEP Math				
Constant	1.99	0.97	2.04	<.0445
OE	0.54	0.12	4.70	<.0001**
Group*	-0.07	0.67	0.11	0.91
Perf Writing				
Constant	1.59	0.16	9.73	<.0001**
OE	0.09	0.02	4.88	<.0001**
Group*	0.44	0.11	3.98	.0001**

* (1=computer)

** statistically significant at the 0.01 level after taking multiple comparisons into account

These results confirm the findings shown in Table 4, namely that even after controlling for OE scores, the effect of mode of administration was highly significant on the PWAvg. However, for the largely multiple-choice NAEP subtests, results indicate no difference for mode of administration.

Performance on Multiple Choice and Short-Answer NAEP Items

Although the regression analysis suggested that mode of administration did not significantly influence performance on the NAEP subtests, further analysis was performed on the NAEP subtest items to examine the effect of administration mode on the two forms of items contained in the NAEP subtest -- multiple-choice and short answer. Table 6 shows the mean score for the two groups on both the multiple-choice items and the short-answer items for the three subtests. Although slight differences between the means were found for the multiple-choice items, none were significant. However, for the science and language arts short answer items, those students who responded on computer performed significantly better than the paper-and-pencil group. While it was expected that performance on multiple-choice items would not differ, the differences detected on the short answer items suggest that even for items that require a brief written response, the mode of administration may affect a student's performance.

The question arises as to why the mode of administration affected performance on the short answer Language Arts and Science questions, but not on the one short-answer Math item. It is likely that the nature of the open-ended Math item accounts for similar performance between the two groups. The open-ended Math question required a short answer which could not be provided without correctly answering the multiple-choice question that preceeded it. In contrast, the three short answer Science items asked students to interpret data in a table, explain their process and respond to a factual item. In particular, the second short answer Science item provided a fair amount of space for a response and many students wrote at least one complete sentence. Although the three Science items were related to the same set of data displayed in a table, response to these items were not dependent on answers to previous items.

Table 6: Results of Analysis of NAEP Subtest Item formats: Multiple-choice versus Short Answer

Items	n	Control Mean	SD	n	Experimental Mean	SD	Effect Size	t	Sig
Lang. Arts									
Mult. Choice	42	8.6	3.47	44	9.0	3.03	0.12	0.64	.522
Short Answer	42	0.6	0.73	44	1.4	0.75	0.99	4.52	<.0001**
Science									
Mult. Choice	42	8.0	3.97	44	9.0	3.99	0.26	1.22	.226
Short Answer	42	0.7	0.77	44	1.7	0.98	1.25	5.06	<.0001**
Math									
Mult. Choice	42	5.8	3.07	44	6.1	3.33	0.10	0.44	.660
Short Answer	42	0.2	0.41	44	0.3	0.47	0.25	1.08	.282

** statistically significant at the 0.01 level after taking multiple comparisons into account

To inquire further into the apparent effect of mode of administration on short answer Language Arts and Science items, we conducted regression analyses, using OE scores as a covariate. Results, shown in Table 7 indicate that the mode of administration had a significant effect on the students' performances on the NAEP Language Arts and Science short-answer items.

Table 7: Results of Regression Analyses on NAEP Language Arts and Science Short-Answer Items

Dependent Var	Coef.	s.e.	beta	s.e.	t-ratio	Sig
NAEP Lang Arts						
Constant	-0.08	0.22			-.38	.71
OE	0.10	0.03	0.35	0.09	3.77	.0003**
Group*	0.63	0.15	0.39	0.09	4.23	.0001**
NAEP Science						
Constant	0.20	0.28			0.74	.4645
OE	0.07	0.03	0.20	0.09	2.09	.0397
Group*	0.91	0.19	0.45	0.09	4.77	<.0001**

* (1=computer)

** statistically significant at the 0.01 level after taking multiple comparisons into account

Discussion

The experiment described here was a small inquiry aimed at investigating a particular question. Motivated by a question as to whether or not performance on an extended writing task might be better if students were allowed to write on computer rather than on paper, the study aimed at estimating the effects of mode of administration on test results for two kinds of assessments, namely the largely multiple-choice NAEP subtests and the extended writing task previously described. Unlike most previous research on the effects of computer administered tests, which has focused on multiple-choice tests and has generally found no or small differences due to mode of administration, our results indicate substantial effects due to mode of administration. The size of the effects was found to be 0.94 on the extended writing task and .99 and 1.25 for the NAEP language arts and science short answer items. Effect sizes of this magnitude are unusually large and of sufficient size to be of not just statistical, but also practical significance (Cohen, 1977; Wolf, 1986). An effect size of 0.94, for example, implies that the score for the average student in the experimental group exceeds that of 83 percent of the students in the control group.

A number of authors have noted the difficulty of interpreting the practical significance of effect sizes and have suggested that one useful way of doing so is with a "binomial effect size display" showing proportions of success and failure under experimental and control conditions (Hedges & Olkin, 1985; Rosenthal & Rubin, 1982). While there are a number of ways in which effect sizes, expressed as either Glass's delta or a correlation coefficient, can be converted to a binomial effect size display, in the case of our PWAvg scores, we have a direct way of showing such a display. Recall that student responses to the performance writing item were scored on a 4-point scale in which scores of 1 and 2 represented a less than adequate response and scores of 3 and 4 represented an adequate or better response. Using the cut-point of 2.5 as distinguishing between inadequate (failure) and adequate (success) responses in terms of PWAvg scores, we may display results as shown in Table 8.

Table 8: Binomial Effect Size Display of Experimental Results: In Terms of Inadequate vs. Adequate PWAvg Scores

Control (Paper)	Inadequate	Adequate
N	32	14
Percent	69.6%	30.4%
Experimental (Computer)	Inadequate	Adequate
N	13	27
Percent	32.5%	67.5%

This display indicates that the computer mode of administration had the effect of increasing the success rate on the performance writing item (as judged by the average of three independent raters) from around 30% to close to 70%.

As a means of inquiring further into the source of this large effect, we conducted a variety of analyses to explore why and for whom the mode of administration effect occurred. To explore why the mode of administration effect may have occurred, we first undertook a textual analysis of student responses to the extended writing task. Specifically we calculated the average number of words and paragraphs contained in the responses of both groups. As Table 9 below indicates, those students who performed the assessment on the computer tended to write almost twice as much and were more apt to organize their responses into more paragraphs.

Table 9: Characters, Words and Paragraphs on Performance Writing Task by Mode of Administration

	Characters	Words	Paragraphs
Control (Paper)			
Mean	586.9	111.6	1.457
Std	275.58	52.47	1.069
n	46	46	46
Experimental (Computer)			
Mean	1022.2	204.7	2.625
Std	549.55	111.32	2.306
n	40	40	40
observed t with pooled variance	4.73	5.07	3.08
sig	<.0001**	<.0001**	<.0001**

** statistically significant at the 0.01 level after taking multiple comparisons into account

In some ways, this pattern is consistent with the findings of Daiute (1985) and Morocco and Neuman (1986), who have shown that teaching writing with word processors tends to lead students to write more and to revise more than when they write with paper-and-pencil. Not surprisingly, the length of students' written responses (in terms of numbers of characters and words correlated significantly with PWAvg scores, 0.63 in both cases). Although this suggests that longer responses tended to receive higher scores, the fact that length of response explains less than half of the variance in PWAvg scores suggests that rated quality is not attributable simply to length of response.

Second, we considered the possibility that motivation might help explain the mode of administration effect. This possibility was suggested to us by spontaneous comments made by students after the testing. For example, after taking the writing assessment on computer, one student commented, "I thought we were going to be taking a test." In contrast, a student in the control group, who had not taken any tests via computer, inquired of us, "How come we didn't get to take the test on computer?" Such comments raised the possibility that motivation and the simple novelty of taking tests on computer might explain the mode of administration effect we found.

Two lines of thought suggest that simple motivation cannot explain our results. If differential motivation arising from the novelty of taking tests on computer was the main cause of our results, it is hard to explain why mode of administration effects were absent on the multiple-choice NAEP subtests, but were prevalent on the performance writing test and the NAEP open-ended items. Furthermore, recent research on the effects of motivation on test performance, suggests that the effects of motivation are not nearly as large as the mode of administration effect we found on the performance writing test. Recently, Kiplinger & Linn (1996) reported on the effects of an experiment in which "low- stakes" NAEP items were embedded in a "high stakes" state testing program in Georgia. Though results from this experiment were mixed, the largest effects of "high stakes" motivation occurred for nine NAEP items designed for eighth grade students. For these nine items, however, the effect size was only 0.18. (Kiplinger & Linn, 1996, p.124). In a separate study, O'Neill, Sugrue & Baker (1996) investigated the effects of monetary and other incentives on the performance of eighth grade students on NAEP items. Again, though effects of these motivational conditions were mixed, the largest influence of motivation ranged from an effect size of 0.16 to 0.24 (O'Neill, Sugrue & Baker, 1996, p. 147). With the largest effects of motivation on eighth grade students found to be in the range of 0.16 to 0.24, these results suggest that motivation alone cannot explain the magnitude of mode of administration effects we found for written responses.

To examine for whom the mode of administration effects occurred, we also inquired into whether the mode of administration effect appeared to be different for different students. First we inquired into whether the mode of administration effect seemed to be different for students performing at different levels on the OE test. One simple way of testing this possibility was to calculate PWAvg scores predicted on the basis of OE scores and see if there was a statistically significant correlation between residuals (actual minus predicted PWAvg scores) and OE scores among the experimental group students. No significant correlation was found, suggesting that the mode of administration effect was not different for students of different ability levels as indicated by their OE scores. A graphical presentation of this pattern is shown in Figure 1, which depicts the line of PWAvg scores regressed on OE scores, with the experimental cases represented with X's and the control group with dots. As can be seen in Figure 1, the actual PWAvg scores for the experimental group tended to exceed the predicted scores across ability levels as represented by the OE scores.

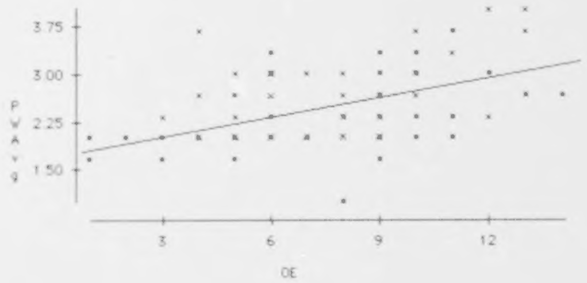


Figure 1: Regression of PWAvg scores on OE Scores

Finally, we explored whether mode of administration effect seemed to differ for males versus females. Table 10 shows PWAvg scores by gender for both control and experimental groups.

Table 10: PWAvg Scores by Gender and Group

	Female	Male	Total
Control (Paper)			
Mean	2.33	2.27	2.30
SD	0.58	0.47	0.55
n	21	25	46
Experimental (Computer)			
Mean	2.92	2.60	2.81
SD	0.53	0.63	0.59
n	26	14	40
Total			
Mean	2.66	2.38	2.53
SD	0.65	0.54	0.62
n	46	39	86

Within the control groups, females performed only slightly better on PWAvg scores than did males (with means 2.33 and 2.27 respectively). However within the experimental group females scored considerably better than males (with means of 2.92 and 2.60). Thus it appears that the effect of computer administration may have been somewhat larger for females than for males. Nonetheless the males who took the extended writing task on computer still performed considerably better than the females who took the writing task on paper (with respective means of 2.60 and 2.33). A two way analysis of variance (PWAvg by gender and group) showed group but not gender to be significant (this was the case whether or not an interaction term was included). This general pattern was confirmed by regression analyses of PWAvg scores on OE scores, sex and group. Though OE scores and the group variable were significant, the sex variable was not.

We should note that post hoc, we were surprised that the proportion of males in the control group (54%) differed by nearly 19 percentage points from the proportion of males in the experimental group (35%). Although the two groups were selected randomly, the probability that this difference would occur is less than .08. However, as can be calculated based on the data in Table 10, even after controlling for gender, the average effect size is 0.86.

Although the experiment reported here had several weaknesses--only one extended writing task was used, no other variables on academic achievement beyond the OE test results were used as covariates in regression analyses, and information on students' extent of experience working on computers was not collected--further research into this topic clearly is warranted.

Increasingly, schools are encouraging students to use computers in their writing. As a result, it is likely that increasing numbers of students are growing accustomed to writing on computers. Nevertheless, large scale assessments of writing, at state, national and even international levels, are attempting to estimate students' writing skills by having them use paper-and-pencil. Our results, if generalizable, suggest that for students accustomed to writing on computer for only a year or two, such estimates of student writing abilities based on responses written by hand may be substantial underestimates of their abilities to write when using a computer.

This suggests that we should exercise considerable caution in making inferences about student abilities based on paper-and-pencil/handwritten tests as students gain more familiarity with writing via computer. And more generally it suggests an important lesson about test validity. Validity of assessment needs to be considered not simply with respect to the content of instruction, but also with respect to the medium of instruction. As more and more students in schools and colleges do their work with spreadsheets and word processors, the traditional paper-and- pencil modes of assessment may fail to measure what they have learned.

We suspect that it will be some years before schools generally, much less large scale state, national or international assessment programs, develop the capacity to administer wide-ranging assessments via computer. In the meantime, we should be extremely cautious about drawing inferences about student abilities when the media of assessment do not parallel those of instruction and learning.

Acknowledgment

We wish to acknowledge our great appreciation of Carol Shilinsky, Principal, and the teachers of the ALL School of Worcester, MA, who first suggested the idea of the study reported here. In particular we thank Rich Tamalavich and Deena Kelly who helped arrange the equipment needed and oversaw administration of the computerized testing. We also wish to thank five anonymous EPAA reviewers who, through editor Gene Glass, provided us with some very helpful suggestions on a previous version of this manuscript with a rapidity that was absolutely astonishing. Five reviews were received within one week of submission of the manuscript! We also thank Gene Glass for suggesting that we take better advantage of the electronic medium via which this journal is published, for example, by appending to this article the full data set for the study so that others might conduct secondary analyses, and by providing electronic links to institutions mentioned in the study. So now, the reviewer who wondered about the unusual school in which this study was conducted can visit the ALL Schools WWW site. Any such visitors will find that students at this remarkable school are now not just writing via computer, but also publishing via the WWW.

References

Barton, P. E. & Coley R. J. (1994) Testing in America's schools. Princeton, NJ Educational Testing Service Policy Information Center.

Beaton, A. E. & Zwick, R. (1990). The Effect of Changes in the National Assessment: Disentangling the NAEP 1985- 86 Reading Anomaly. Princeton, NJ: Educational Testing Service, ETS.

Bunderson, C. V., Inouye, D. K. & Olsen, J. B. (1989). The four generations of computerized educational measurement. In Linn, R. L., *Educational Measurement (3rd Ed)*. Washington, D.C.: American Council on Education, pp. 367-407.

Cizek, G. J. (1991). The Effect of Altering the Position of Options in a multiple-choice Examination. Paper presented at NCME, April 1991. (ERIC)

Cohen, D. (1990). Reshaping the Standards Agenda: From an Australian's Perspective of Curriculum and Assessment. In P. Broadfoot, R. Murphy & H. Torrance (Eds.), *Changing Educational Assessment: International Perspectives and Trends*. London: Routledge.

Cohen, J (1977). *Statistical power analysis for the behavioral sciences* (rev. ed.) NY: Academic Press.

Daiute, C. (1985). *Writing and computers*. Reading, MA: Addison-Wesley Publishing Co.

Darling-Hammond, L., Acness, J. & Falk, B. (1995). *Authentic Assessment in Action*. New York, NY: Teachers College Press.

Dunbar, S. B., Koretz, D. M., & Hoover, H. D. (1991). Quality Control in the Development and Use of Performance Assessments. *Applied Measurement in Education*, 4(4), 289-303.

Green, B. F., Jr. (1970). Comments on tailored testing. In W. H. Holtzman (Ed.), *Computer-assisted instruction, testing and guidance*. New York: Harper and Row.

Haas, C. & Hayes, J. R. (1986). What Did I Just Say? Reading Problems in Writing with the Machine. *Research in the Teaching of English*, 20(1), 22-35.

Hancock, G. R. & Klockars, A. J. (1996). The quest for ∝: Developments in multiple comparison procedures in the quarter century since Games (1971). *Review of Educational Research*, 66(3), 269-306.

Hedges, L. V. & Olkin, I. (1985) *Statistical methods for meta-analysis*. San Diego: Academic Press.

Kiplinger, V. L. & Linn, R. L. (1996). Raising the stakes of test administration: The impact on student performance on the National Assessment of Educational Progress. *Educational Assessment*, 3(2), 111-133.

Mead, A. D & Drasgow, F. (1993) Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta- analysis. *Psychological Bulletin*, 114(3), 449-458.

Morocco, C. C. & Neuman, S. B. (1986). Word processors and the acquisition of writing strategies. *Journal of Learning Disabilities* 19(4), 243-248.

Mourant, R. R, Lakshmanan, R. & Chantadisai, R. (1981). Visual Fatigue and Cathode Ray Rube Display Terminals. *Human Factors*, 23(5), 529-540.

O'Neil, H. F. Jr., Sugrue, B. & Baker, E. L. (1996). Effects of motivational interventions on the National Assessment of Educational Progress mathematics performance. *Educational Assessment*, 3(2), 135-157.

Rosenthal, R. (1994) Parametric measures of effect size. In Cooper, H. & Hedges, L. *The handbook of research synthesis*. NY: Russell SAGE, pp. 231-244

Rosenthal, R. & Rubin, D. B. (1982) A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology*, 74, 166-169.

Snyder, T. D. & Hoffman, C. M. (1990). Digest of Education Statistics. Washington, DC: U. S. Department of Education.

Snyder, T. D. & Hoffman, C. M. (1993). Digest of Education Statistics. Washington, DC: U. S. Department of Education.

Snyder, T. D. & Hoffman, C. M. (1994). Digest of Education Statistics. Washington, DC: U. S. Department of Education.

Snyder, T. D. & Hoffman, C. M. (1995). Digest of Education Statistics. Washington, DC: U. S. Department of Education.

Wolf, F. (1986) *Meta-analysis: Quantitative methods for research synthesis*. SAGE University series on quantitative applications in the social sciences, series no. 07-059. Newbury Park, CA: SAGE.

About the Authors

Michael Russell

russelmh@bc.edu

Boston College
School of Education
Center for the Study of Testing, Evaluation and Educational Policy
323 Campion Hall
Chestnut Hill, MA 02167
Ph. 617/552-4521
Fax 617-552-8419

Walt Haney

haney@bc.edu
Boston College
School of Education
Center for the Study of Testing, Evaluation and Educational Policy
323 Campion Hall
Chestnut Hill, MA 02167

Copyright 1997 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://olam.ed.asu.edu/epaa

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey@olam.ed.asu.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Marshall University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Rocky Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
Arizona State University

Dennis Sayers
University of California at Davis

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina--Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher Education

Mary P. McKeown
Arizona Board of Regents

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois--UC

Robert T. Stout
Arizona State University

---

archives | abstracts | editors | board | submit | comment | subscribe | search

EPAA • enter the archives • browse the abstracts • the editors • the edit board
• submit article • submit commentary • search • subscribe
volume: • 1 • 2 • 3 • 4 • 5 • 6 • 7

This article has been retrieved 1232 times since January 23, 1997.

Education Policy Analysis Archives

Volume 5 Number 4 **January 23, 1997** **ISSN 1068-2341**

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass Glass@ASU.EDU.

College of Education

Arizona State University,Tempe AZ 85287-2411

Copyright 1997, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

Learning from Others:
Service-Learning in Costa Rica and Indonesia

David D. Williams
Brigham Young University

William D. Eiserman
Colorado Foundation for Families and Children

Abstract:
Calls are increasingly sounded for universities to better address their communities' and students' needs through service, as well as research and teaching. This article invites policy makers to re-examine university service, research, and teaching responsibilities by reflecting on roles service-learning plays in universities in Indonesia and Costa Rica. We conclude that service-learning plays a critical role and a key to expanding service-learning for students and understanding the utility of such a policy change is increased faculty involvement. Until more faculty explore the "why" and "how" of service-learning, research and teaching will dominate the university agenda.

> "Martin Luther King Jr. once said that "our scientific power has outrun our spiritual power. We have guided missiles and misguided men." I have seen many times men who abandoned their dreams; I have seen those misguided men who believe they can solve all problems with guided missiles. These are men without values, and the

world cannot afford their leadership into the twenty-first century.

"The return to a life and a world dominated by values is urgent if we want peace to prevail. We should no longer be ashamed of feelings of piety. It is not true that they degrade reason and science. Piety is no less than the intelligence of the soul, and we need heart and brains to recover the world in our hands, for the values we cherish.

"Nobody can ignore the problems of today, least of all the intellectuals." (Arias, 1988, pg. 19)

Oscar Arias Sánchez
President of Costa Rica

A new book on service-learning in higher education (Jacoby, 1996) "provides a historical overview and a context for understanding the essential linkage of service and learning; it describes the current state of practice; and it highlights the relationship between service-learning and institutional educational goals" (p. 5). After examining the predominant assumptions underlying the combining of community service and academic learning in higher education and offering several illustrative examples from colleges and universities in the United States, several authors address in part three of the book "organizational, administrative, and policy issues" which "may be the most crucial factors in the initiation and sustainability of service-learning" (p 229).

Reviews of this excellent summary of the field and several related sources (e.g., Albert, 1994; Daloz, et al., 1996; Kendall and Associates, 1990; and many others which are indexed and available through the University of Colorado's service-learning homepage (http://csf.colorado.edu/sl/) raise some questions for educational policy makers to consider:

- What kinds of service belong in higher education?
- How does service enhance and/or detract from learning, teaching, scholarship, and other institutional goals?
- What policies regarding service should be made in higher education?
- What evidence is accruing that might inform policy regarding potential roles of service-learning in higher education?

Though most universities have always claimed that "service" is one of the three main purposes for higher education, both research and teaching continue to dominate the activities of most academics and their institutions. However, growing numbers of community service proponents are arguing that service combined with other kinds of academic learning should receive a more equitable place in higher education. While traditionalists note that there is very little time, incentive, or support for more service in academic life, service-learning proponents insist that the three-pronged mission need not constitute three separate sets of activities. Rather, they contend that activities under each of these missions may be more effective and efficient when integrated into a common set of activities (Jacoby, 1996).

As questions are asked and plans are developed for service-learning programs, whether on individual, institutional or national levels, it would be wise to learn from the experiences of those who have been involved in the development and implementation of service-learning programs in various contexts. Existing programs may inform a new "vision" for higher education's service role and may shape the development of practices for fulfilling that "vision" as well.

Unknown to many higher educators, some of the most comprehensive and innovative approaches to service-learning have been designed and implemented in developing countries (Eberly & Sherraden, 1990). In this study, the University of Costa Rica's compulsory service-learning program, which began in 1975, is explored and compared to a similar, even older program in Indonesia to help readers consider some questions to ask as they examine the role of service-learning in higher education in other societies and in their own contexts.

Following a short case study of a service-learning project in San José, Costa Rica, a brief definition of service-learning, a summary of principles identified in the literature which should under gird such projects, and an overview of methods used in our inquiry, this article explores what can be learned by educators world-wide through understanding the historical roots, program components, perceived outcomes, and perceived strengths and weaknesses of the two programs in Costa Rica and Indonesia.

Helping Children Help Children: One Brief Case

Entering a slum community set aside by the government of Costa Rica for indigent families who own no property, our van pulls off the blacktop onto a dirt road strewn with garbage and dirty-faced children in tattered clothes. Discarded plastic bags, old bottles, and cans clog an open sewage ditch which reeks of stagnant human waste and rotting trash. Stopping in front of a plywood and tin shed, Marta Picado Mesen, a Social Work professor from the University of Costa Rica, explains that this shed was built by World Vision and is intended to be used as a meeting hall exclusively by people living in one of the nine sectors of this settlement. We are visiting this settlement with Marta and four of the twenty-two university students who have joined her for the past few months. This settlement is the focus of a project they are doing as part of the university's community service program, Trabajo Comunal Universitario (TCU), in which students are required to participate before graduation.

As we unload the van, Marta explains that despite the perceptions of many tourists that Costa Rica is a clean and safe, idyllic place, there are many problems with crime, health, and drug abuse. This settlement is a sort of breeding ground for the worst of such things. Brought here from all over Costa Rica, many of these people were removed from squatter sites, while others had been homeless. Now they live here, on the outskirts of Costa Rica's capitol, San José, having been placed in the particular sector of the settlement which corresponds with the section of the country from which they were removed. The majority of the 3,840 individuals living here are members of single parent families and earn no salary. Forty-eight percent of the settlement's residents are children under the age of fifteen and the average family income is between 5,000 and 22,000 colones per month (approximately $36 to $160)--well below the poverty line in Costa Rica. About 15% of the inhabitants are unregistered Nicaraguan refugees, though it appears that nearly everyone in this settlement is a refugee of sorts.

Nearby, men are loudly nailing a sheet of rusted, corrugated tin to a small frame hut to make walls and a roof; a new home in the making. Eager to share their experience, the students explain that people move in every day and are constantly searching for materials to build shelters for themselves. Nearly everyone lives in multiple family dwellings. Each of the twelve-square-meter shelters, of which only 60% have latrines, is home to about 16 people. The only public building is a shed-like school on the edge of the settlement which is staffed by five teachers who teach about 500 children a day in shifts from six in the morning to six at night.

Surprised by the severity of these circumstances, we would like to wander through the settlement to see all of this more close-up, but Marta warns us that it is too dangerous. We must stay here on the edge of the settlement in the World Vision building, with the van parked outside. Inside the tin-roofed meeting hall with a dirt floor, while we sit at weather-beaten tables, Marta explains that the 960 families which live in the nine sectors of the settlement fight among themselves, with gang leaders from each sector leading assaults against people in other sectors. Children are often caught in the middle and, Marta believes, are consequently at serious risk psychologically, physically, and educationally. Marta's TCU project aims to address the needs of some of these children.

The focus of her project is to help a selection of 30 children in the settlement prepare themselves to more effectively deal with the problems they face due to the conditions in which they live and to reach out as peer leaders to help other children. The scope of the project is broad, starting with activities associated with helping the children understand the dangers of drug abuse, the importance of education, the need to obtain health care, and the need to prevent disease, physical abuse, prostitution, and sexual promiscuity. As the project proceeds, Marta expects it will become more focused on concerns and interests of the participating children. They have already begun to elicit information from the children which will direct the planning for the next phase of the project.

While Marta serves as the director of the project, most of the direct work with children and other members of the community is being done by her TCU students who come from a wide variety of disciplinary backgrounds including nursing, medicine, social work, psychology, dramatic arts, and education. Marta has given four of the students (who are from health, psychology, dramatic arts, and education) responsibility for presenting the project to us. Prepared with handouts and using a portable overhead projector plugged into an outlet from a dangling ceiling light bulb, one-by-one the students present different aspects of the project including the settlement statistics, an overall project description, their specific objectives, the problems they have encountered, and their accomplishments.

As they present, a dozen or more children from the settlement playfully throw rocks on the roof of our building. Though these rocks are distracting to us as they thunderously roll off the tin roof, this interruption scarcely interferes with the students' steady, enthusiastic presentation of their experiences. These childish pranks, which appear to be commonplace to the students, remind us throughout their presentation of the disturbing harshness of the circumstances in which Marta and these students have been working.

Devoting their entire 300 hours of community service required for graduation from the University of Costa Rica on the initial diagnostic and planning phases of this project, these students have met with formal and informal leaders of the settlement to explain the project and to coordinate their efforts with other organizations within each sector of the settlement. From these efforts, they identified thirty children between the ages of eight and twelve from each of the sectors who are participating in the project. The students then diagnosed the children's challenges with respect to health threats, social problems, and educational needs through data gathering activities including games, discussions, and role-playing directly involving the 30 children. With their 300 hours now complete, these TCU students will be passing the project on to a new group of students who will continue to develop and carry out the next phase of the project, with Marta.

While they have confronted numerous challenges along the way, Marta and the students identify two areas with which they have been especially concerned. First, the students explained that they and the children are from completely different worlds. Previously they had had very little understanding of what these people faced economically and socially. The students consequently question the extent to which they can appropriately reach out to the children with what the children most need. A similar lack of understanding is true in reverse. As one student expressed,

> We know we are looked to as role models by some of these children, but what that means since we are from different worlds, I don't know. Is that helpful? Most of them will never have the opportunities we have. (Note 1)

Second, as this first phase is ending, there is going to be a total team turnover, with the exception of Marta. This represents a challenge since the children's participation in the project is largely a function of the relationships they develop with the TCU students. According to Marta, it is also somewhat traumatic for the TCU students since they come to know and care about these children:

> Making this transition and maintaining continuity within the project is the challenge we face next.

Marta recognizes that they face many challenges in being able to effectively empower children to help other children face some tremendous difficulties of life in this situation. The importance of this project, according to Marta, is that they are addressing concerns for which there presently are no obvious solutions. In this respect, the TCU project provides a forum for Marta's scholarly research as well, including exploration of such questions as: Will the concept of children helping children work under these conditions, within the culture of this settlement? What will those who are assisted in their peer leadership abilities gain from trying to help other children? Marta explains,

> These are some of our questions. We will have to see. We know that they need what we are trying to provide. We just have to try.

While the focus of the experience has clearly been on what could be provided to the children of this settlement, the university students say this has been a valuable learning experience for them as well. They each describe ways in which they have creatively applied their education to the problems in this situation. For example, the drama student explains how rewarding it has been to use role playing to help the children identify their social problems. He had never seen his art form in this way; actually applied in such a useful manner.

The health student explains that in his previous coursework he had learned about the existence of various health problems and difficulties associated with getting people such as these to use free health services to which they were entitled. His work in this project has given him a chance to actually be a part of an effort aimed at overcoming some of those problems; according to him, a "way" of learning about these issues which other coursework could not have provided.

All of the students say that beyond gaining an appreciation of life in these harsh conditions and an understanding of how their skills could be useful there, both of which address the primary aim of the TCU program, they have learned the value of combining their disciplines with others in a collaborative effort. According to these students, interdisciplinary problem solving in actual, real- life settings is often what distinguishes TCU learning from other coursework.

Three of the four students believe they probably would have participated in some form of service even if TCU had not been compulsory, but they do not know how they could have organized anything like this on their own. Only one student says she would not have been involved in a service project like this if it had not been required because her family did not like the idea of her going into such a dangerous place.

As we leave the settlement with Marta and her students we are filled with questions. Can they succeed in helping these children? How? Is this what is meant by the integration of research, teaching, and service? Is this really working as well as it already seems to be? Why here? Why this project when there are so many other things they could all be doing which would more likely succeed and which would be easier and safer? Why does Marta feel that they "just have to try"?

Principles

Service-learning has been defined in many different ways but the definition used by Jacoby (1996, p. 5) is informative:

> Service-learning is a form of experiential education in which students engage in activities that address human and community needs together with structured opportunities intentionally designed to promote student learning and development. Reflection and reciprocity are key concepts of service-learning.

Jacoby argues persuasively that service-learning brings the resources of the university to bear on social issues of concern which are not adequately addressed by other means. In addition to benefits for students, faculty, and the community, service-learning may help upgrade the educational institution itself:

> Higher education is being called on to renew its historic commitment to service. Its foremost experts are urging colleges and universities to assume a leadership role in addressing societyUs increasing problems and in meeting growing human needs.... At the same time, higher education is questioning its effectiveness at achieving its most fundamental goal: student learning.... As colleges and universities across the country are developing programs to enable their students to serve their communities, the nation, and the world-- and at the same time to enrich undergraduate education-- it is critical that these programs embrace the concept of service-learning. (Jacoby, pp. 3-5)

A growing number of university policy shapers are accepting the claim that the traditional curriculum and research agenda of universities can be informed via service- learning activities; and community support to the institution may be enhanced in many different ways as a result of the heightened visibility of the institution through its community-based service-learning activities.

While its proponents suggest that service-learning has the potential of addressing a wide variety of aims, perhaps most emphasized is the rationale of service-learning as a viable, maybe even optimal means for impacting students and faculty with regard to the improvement of social and civic responsibility, enhanced intellectual development, cross- cultural learning, leadership development, moral and ethical development, and career development (Kendall and Associates, 1990). Most service-learning programs tend to focus on one or two of these areas rather than address all of them equally, resulting in a wide array of service-learning configurations.

To understand the common threads of service- learning programs, as well as to help construct service- learning programs which possess the most critical elements, in 1989 a set of ten principles was developed by a group of service-learning educators from across the United States. These principles were intended to represent the common threads which distinguish service-learning from other types of learning and from other types of service activities. These principles state that an effective service-learning program:

1. Engages people in responsible and challenging actions for the common good.
2. Provides structured opportunities for people to reflect critically on their service experience.
3. Articulates clear service and learning goals for everyone involved.
4. Allows for those with needs to define those needs.
5. Clarifies the responsibilities of each person and organization involved.
6. Matches service providers and service needs through a process that recognizes changing circumstances.
7. Expects genuine, active, and sustained organizational commitment.
8. Includes training, supervision, monitoring, support, recognition, and evaluation to meet service and learning goals.
9. Insures that the time commitment for service and learning is flexible, appropriate, and in the best interests of all involved.
10. Is committed to program participation by and with diverse populations (Kendall and Associates, 1990, p. 40).

While there seems to be considerable agreement among service-learning educators regarding these principles generally, discussion abounds on a number of issues related to how these principles are to be expressed in practice. In response, we have conducted two studies in universities in developing countries (the Kuliah Kerja Nyata or KKN projects through the

University of Indonesia and the Trabajo Comunal Universitario or TCU projects through the University of Costa Rica) which have been practicing diverse versions of service- learning for many years, to better understand how their practices relate to these principles and to discover other principles that might be helpful in guiding similar efforts elsewhere.

Methods

To begin exploring these issues, a qualitative inquiry (Denzin & Lincoln, 1994) approach was used, allowing for emergent themes to arise in the context of on-site interviews and observations. Qualitative inquiry provides a means for investigators to refine their questions to better reflect the perspectives of all participants throughout a study. Thus, an attempt was made to blend the concerns of students, faculty, administrators, service recipients, the literature, and the researchers through ongoing refinement of questions in light of concurrent data analyses. New questions arose and were addressed along with questions suggested by the literature.

The second author reviewed documents and interviewed participants in three KKN projects associated with Andalas University in Padang, West Sumatra, Indonesia. Then, after analysis and review of that experience, both authors visited eleven TCU projects associated with the University of Costa Rica (of which the case study presented earlier was one) throughout Costa Rica. Details regarding data sources for both studies are summarized in Table 1.

Table 1
Summary of Data Sources for the Indonesian and Costa Rican Studies

Data sources	**Indonesia (KNN)**	**Costa Rica (TCU)**
Administrators interviewed	3	10
Faculty/supervisors interviewed	1	12
Students interviewed	19	11
Student questionnaire	19	NA
Community members interviewed	15	30
Projects studied	3	11

Interviews, in Indonesian and Spanish, were conducted with administrative staff, participating faculty (from accounting/business, agriculture, animal husbandry, arts and letters, civil engineering, chemistry, law, and medicine in Indonesia and from history, social work, engineering, linguistics, agronomy, art, anthropology, education, computer science, and nursing in Costa Rica), participating students (from accounting/business, agriculture, animal husbandry, arts and letters, civil engineering, chemistry, law, and medicine in Indonesia and from geology, psychology, drama, English teaching, nursing, botany, and nursing in Costa Rica; many other fields such as architecture, social work, education, art, engineering, sociology, and so on were involved in the Costa Rican projects but we were not able to meet with students from all represented disciplines), as well as various community participants (families being served, teachers in the settings, and visitors at museum and display sites).

We were able to travel to several project sites to see either the results of the projects or to see them in progress. These visits allowed us to observe KKN and TCU participants at work, gave us a sense of the outcomes from their efforts, and provided us with general observations of the contexts in which projects were implemented.

Three forms of data were compiled in this inquiry. The primary data were field notes taken by the two investigators. These field notes included reconstruction of interviews and observations as well as our questions and interpretive comments. Supplementary data included photographic documentation of sites, observable outcomes of projects, and exposure to the various participants. Additionally, archival data were collected which were relevant to the questions addressed in this investigation, including official documents describing the goals and objectives of the programs and their implementation criteria, documents provided to faculty and students regarding participation guidelines, student evaluation instruments, as well as student and faculty reports of individual project activities.

Analysis procedures consisted of three activities:

- First, key questions to guide the inquiry were identified through a review of literature and our own experiences with service and learning.
- Second, we elaborated and expanded the key questions by reflecting on information obtained throughout each country's study.
- Third, we interpreted our experiences in these sites by searching for patterns across data sources and by attempting tentative answers to both our original guiding questions and to questions which emerged throughout the study.

Several methodological standards for conducting qualitative inquiry have been proposed (e.g., Eisner, 1991; Guba and Lincoln, 1989) and were used to guide this study. Though each visit was brief (less than a month in each country) we were able to meet the triangulation standards by using multiple investigators, sites, informants, and collection procedures. We also shared our findings with participants and asked for their judgments of accuracy and credibility (member checking), shared our findings with disinterested others to discover our blind spots (peer debriefing), and searched for evidence that would counter our conclusions (negative case analysis) to ensure trustworthiness of the findings. Finally, we have included a case description of one project (at the beginning of this article) to allow readers to hear the voices of the participants and to judge transferability of our findings to readers' sites. We have also kept an audit trail of all our activities throughout this project to increase the likelihood of dependability and confirmability of the study.

Lessons Learned

So what did we learn from participants in these two countries that could help others as they contemplate service- learning and policy setting in university settings? In the remainder of this article, we summarize the lessons learned around the following questions, which are a combination of questions we began the study asking and other questions that arose:

1. What are the historical roots that lead to the formation of these programs?
2. What are the basic program purposes and components?
3. What are the perceived outcomes, concerns, and lingering questions associated with participation in these programs from various perspectives?
4. What are some implications for combining service and learning for university students in other countries?
5. What questions are raised by this study for future inquiry into service and learning in higher education?

Historical Roots

Both of these "non-military national service" programs developed in the 1970's, beginning in grassroots initiatives and culminating in centralized governmental support and/or mandates.

Indonesia. Rooted in a rich history of gotong-royong (or mutual assistance), KKN had its first observable roots as a program implemented between 1945 and 1949 when Indonesians were struggling for independence from the Dutch. Due to a critical lack of teachers in guerrilla areas at that time, members of the student army were recruited to teach in secondary schools in these areas (Hardjasoemantri, 1981). After the fighting ceased, the organized students felt a "moral commitment" to continue providing teaching services on a voluntary basis and subsequently developed a volunteer project called Pergerahan dan Penempatan Tenagaa Mahasiswa (recruitment and placement of students for the purpose of teaching, known as PTM) which lasted until 1962.

In 1966 a major revision of the entire educational system included an institutional service-learning concept that was drawn from the earlier students' experiences with PTM. Several reform objectives eventually emerged from this movement which would become the basis for Kuliah Kerja Nyata or KKN:

1. education would become more Indonesia-based in content,
2. education would relate more closely to the range of skills presently needed in Indonesia,
3. the availability of non-formal education would be increased in order to complement the available formal education, and
4. education would provide greater opportunities for young Indonesians to participate directly in the development of their country in practical and satisfying ways.

With these objectives as a foundation, KKN emerged in 1972 as a formal course which was piloted at three of Indonesia's major universities, including Andalas University. The success of these pilot projects resulted in expansion to all 43 public universities, 90% of which subsequently have determined that KKN is compulsory for all students. Thus the impetus behind service-learning has shifted over the years from a voluntary effort fueled by student initiative to a compulsory "program" mandated by universities.

Costa Rica. Details from Sherraden & Castillo (1990), Gonzalez (1992), a booklet describing TCU entitled Información General (1992) and interviews with TCU administrators and university faculty contribute to understanding the historical roots of TCU as a program initiated by several Costa Rican students and faculty concerned about how to make their university more responsive to the needs of their broader society. They felt that because they were receiving many opportunities at the expense of others, they ought to find a way to compensate the rest of the nation for what they had received. This attitude first emerged in what is described as a small "radically left" group of students and received support from a similar group of faculty during the 1960's and 1970's. Like most modern universities, the University of Costa Rica identifies a three-pronged mission for itself, including teaching, research, and service to the community.

Therefore, in 1974, in an effort to respond to this "service movement," to raise service to the level of importance held by research and teaching, and to better integrate the three within the academic mission of the institution, the University of Costa Rica created an Office of Social Action (Vicerrectoria de Acción Social) to match its two sibling offices of research and teaching; all three of which serve as administrative supports for faculty. Responsible for a variety of service related activities which bridge faculty and students with the community, this office administers several service-learning programs (including TCU which became compulsory for students in 1975) which involve faculty and students from each academic department to promote service as an integral part of the academic mission of the university.

Basic Program Purposes and Components

Costa Rica and Indonesia have many similar objectives and processes for carrying out their service-learning programs. These are summarized in Table 2.

As the details in Table 2 suggest, there are many similarities between these two programs but there are also tremendous differences. In terms of purposes, both programs are viewed as means for bridging university resources with community resources. The emphasis in both is on community improvements and benefits while benefits to the students, faculty, and university as separate from the community are of secondary importance. They both emphasize students' obligations to society more than student learning, although that is an obvious secondary focus.

Table 2
Program Purposes and Components of the Indonesian and Costa Rican Programs

	Indonesia	Costa Rica

Purposes	(For Indonesia, these are adapted from Direktorate Pembinaan dan Pengabdian pada Maysarakat Ditjen Kikti Depdikbud, 1986.) 1. Students obtain learning experiences through their involvement in the social life of people in the community where they are directly exposed to everyday problems and address those problems in the process of development pragmatically and inter-disciplinarily. 2. Students contribute their ideas to the people by virtue of sciences, technologies, and arts in the attempt to stimulate and escalate the growth and development of the community as well as set up cadres for continuing development. 3. Universities produce graduates who are more aware of the complex conditions, changes, and problems that the people face in the process of development. Therefore, the graduates of universities can be prepared to overcome problems pragmatically and interdisciplinarily. 4. Relations among universities, local government, technical offices and members of the community are strengthened. Consequently, universities will be able to play a broader role and adapt their educational and research activities to the actual demands of the developing society	(For Costa Rica, these are adapted from Gonzalez, 1992, pg. 8.) 1. To raise the social consciousness of future professionals by bringing them into direct contact with their society and its problems. 2. To partially reimburse the society for what it has invested in the preparation of its university students. 3. To promote the students' sense of social responsibility so that they will continue to serve their communities throughout their professional careers. 4. To provide feedback to the university regarding how well it is meeting its teaching and research missions. The confrontation of the academic world with the social environment in service-learning settings should lead to important changes in curricula and research projects.
Student Motivation	Compulsory for all undergraduates. Generally completed after the third year at the university.	Compulsory for all undergraduates. Generally completed after the third year at the university.
Faculty Motivation	Faculty are assigned to KKN supervisory responsibilities as part of their regular assignment.	Varies. One faculty member (minimum) from each academic department elects to be involved by developing their own project.

Community Motivation	Varies. But usually participation is viewed as a means of obtaining ad-ditional services not otherwise available and sometimes upgrading the status of the community generally.	Varies. But usually participation is viewed as a means of obtaining additional services not otherwise available and sometimes upgrading the status of the community generally.
Program Office	A separate office designed to bridge other units on community and research; reorganization underway to shift KKN under community unit	Equated with sibling offices on research and teaching; provides support to department level TCU activities.
Project development and administration	Centralized within KKN office with guidance from regional and national government input on priorities. Faculty assigned to KKN developed projects and sites.	Decentralized to academic departments and voluntary faculty within those departments. Faculty develop their own projects and get approval by TCU office.
Length of project	Usually two months	Usually three years
Length of service for students	Two months full time-- projects are turned over to the community after this with no new students coming in.	300 hours-- projects usually continue with another group of students and the same faculty member.
Curricular placement	Separate from traditional coursework and concentrated in June-August.	Separate from traditional coursework but ongoing throughout the year.
Preparation of students	Prior to participation in the program, students complete a coaching orientation.	Prior to participation in the program, students complete a large group course on "national reality."
Project choices	Students are assigned to a team which is assigned to a project setting.	Students are provided with a booklet of about 100 faculty projects from which they may choose and apply.
Setting/focus	All projects are similar. They are interdisciplinary projects focusing on several areas of development reflecting regional and national governmental priorities, usually in rural settings but moving toward urban projects in the future.	Varies greatly from foci on cultural development, to health education, to infrastructure development to literacy concerns. May be focused on a single social issue or on multiple aspects of urban or rural concerns.
Team size (on site)	8 to 10 students and a supervisor	Varies: 10 to 30 and a faculty member.
Team composition	Interdisciplinary	Interdisciplinary
Residency	Students are required to reside in the village community where the project takes place.	Students generally continue to reside at home.

Supervision	Varies but faculty serve as distant supports and evaluators. Students work very independently with little supervision.	Varies but tends to have faculty involved in all phases of projects. Students usually have regular supervision.
Community participation	Varies greatly. Mechanisms exist to promote the development of projects which arise from needs identified by community members, which foster community participation.	Varies greatly. Mechanisms exist to promote the development of projects which arise from needs identified by community members, which foster community participation.
Program evaluation	Formal formative and summative evaluations involve all stake holders.	Formal formative and summative evaluations involve all stake holders.
Student evaluation	Graded, group grading used.	Pass/fail
Faculty involvement	A small core group of faculty (about 20% of all faculty) are actively involved.	A small core group of faculty (about 20% of all faculty) are actively involved.
Reflective component	There is no specified reflective component but there are regular problem-solving/decision-making meetings. These may be formal with participation of faculty and community members or informal with only students. Implicit emphasis is to focus on others and not encourage self-conscious reflection on service.	There is no specified reflective component but there are regular problem-solving/decision-making meetings. These may be formal with participation of faculty and community members or informal with only students. Implicit emphasis is to focus on others and not encourage self-conscious reflection on service.
Finances	Minimal funds are used. Supplemental funds are occasionally received from extension offices. Students are provided one round trip transport to and from the site. Most students pay tuition and often contribute project funds.	Minimal funds are used. Supplemental funds are occasionally received from extension offices. Provides food and transportation to and from out of city sites. 70% of students pay no tuition and students rarely fund projects.

In terms of motivations for participating, community members view the programs in both countries as a means of obtaining services not otherwise available. These programs are also meant to provide a feedback loop so that society can inform the university about the social realities that academia should address. Thus, the community not only receives benefits but should also inform the university via these programs.

Students in both countries are required to participate if they want to graduate (though many of them want to offer their help and do not view this requirement negatively). According to faculty and students interviewed in this study, students who participate in Costa Rica feel some obligation, beyond their university requirement, to participate because about 60% of them pay almost no tuition. Even of those who do, the highest payments are only about the equivalent of $120 a semester. And in Indonesia, all university students feel some obligation to the rest of their society to help, even though they have to pay tuition and often major costs associated with their projects.

Faculty motivations vary more substantially. In Indonesia, faculty are assigned supervisory roles to projects that may or may not be directly relevant to their teaching and research agenda. They do not appear to glean much professionally or personally from participating. But in Costa Rica, faculty volunteer to be involved by proposing their own projects which sometimes grow directly out of their research and teaching activities. It appears that the main incentive for professors' involvement in TCU is to write about it. They are welcome to publish what they learn, just as they would with regular research projects.

Administratively, the programs are very similar in that program offices are set up to bridge service with research and teaching rather than make service a separate activity. However, the Indonesian program administration is centralized within the KKN government office and involves service learning projects for students from many different universities. Though the Costa Rican program has the TCU office and a formal project development and evaluation process (involving reviews of plans, implementation, and outcomes), most of the development and administration is decentralized to academic departments and faculty within those departments. Each academic department in the university has a faculty member assigned to coordinate the efforts of their department with the Office of Social Action. This person helps orient other professors to the Social Action programs, knows the community service projects pertaining to their department, helps solve problems encountered by participants in the program, encourages integration of service with inquiry and teaching, and otherwise searches for ways to meet the office of Social Action's main responsibility, which is to translate what the university is learning into the society at-large.

Both programs provide minimal financial support, although the Costa Rican program appears to provide slightly more. Most University of Costa Rica departments dedicate at least three percent of their budget to TCU-related projects and allow up to 30% overload or faculty release time to participate in these programs. The TCU office provides an assistant for ten hours a week to help faculty members in whatever ways they see fit. That office also provides food, transportation, hourly assistants, materials, and evaluation/ accreditation assistance for projects they approve. In contrast, the Indonesian students often provide their own financial support for projects in addition to paying tuition, though the program provides travel to and from the project site.

In terms of student requirements and faculty involvement, the TCU program, translated as university community work, is designed to meet the Costa Rican objectives through compulsory "pass/fail" participation of all undergraduate students, in addition to their traditional coursework and departmental practica requirements. Students must complete the equivalent of 300 hours of service by working on a segment of a faculty member's TCU-approved project in less than one year after completing at least 50% of their coursework and taking a class on "national reality" which orients them to the problems of the nation (course content and activities vary widely as each academic department teaches its own version). They may apply to participate in a project after reviewing available project descriptions in a booklet.

These projects usually involve some needs assessment with community leaders or members, span at least three years and are somewhat interdisciplinary in nature. Project foci vary widely from efforts to solve health problems, to improving literacy, to enhancing cultural development, to preserving historical relics, to preserving native dialects. Projects usually grow out of the faculty members' assessment of what a particular community needs in light of each faculty member's primary research and teaching interests and their ongoing relationships with the community. Thus, the faculty members are usually members of these communities or have strong ties to them and are willing to dedicate several years to addressing needs there, working with several cohorts of 10-30 students from many different disciplines throughout the project's history.

The KKN program is also compulsory and subsequent to at least two years of on-campus study. It involves participation in a graded four credit "coaching" class taught for 2-4 hours from January to June within each college (usually in large groups of up to 200 students) to orient students to the infrastructure of villages, as well as production, education, social, cultural, and spiritual issues important in village life. Toward the end of this course, students are divided into small teams of 8-10 which are assigned to participating villages and they meet to prepare for an initial visit to the village. Meanwhile, village leaders are approached by government and university staff to explain the aims of KKN and prepare the villagers to identify needs and prepare for the team to take residence there. Eventually, an initial one day "observation visit" is conducted to allow students to gather data about the situation and needs of the village. However, rather than a formal needs assessment, this visit usually only involves introductions, a short tour of the village, and a short meeting to discuss the starting date, housing arrangements, and the expression of hopes for what will be accomplished.

About a week after this initial visit, the student team moves into the village and spends two months developing relationships, working with village leaders to develop a multi-disciplinary work plan and working collaboratively with one another and villagers to address this plan. The students are supposed to play five main roles: sharers of information from outside sources that the villagers might want to use, motivators to encourage village members to make necessary changes, diffusers of national programs and ideas, inter-system mediators between villagers and offices offering technical services within the region, and supervisors of project activities. However, actual roles emerge and are negotiated, often resulting in the students spending considerable time physically laboring in the village on projects villagers want. The students also spend time each day informally meeting together to talk about their work, challenges they face, and possible solutions.

The foci of the projects usually combine several areas of development reflecting local, regional, and national governmental priorities. Some examples include rodent control, building an irrigation canal, reactivation of a local chapter of the national family education and welfare organization for women, activation of youth in village projects through sports, traditional dancing, and drama, renovation of a bridge, advice on legal issues, conservation of traditional folk drama through education of the youth by knowledgeable villagers, education on health maintenance and hygiene, conducting a census, and mobilizing funds for economic development. The students are supervised and evaluated/graded at the end of the two months by faculty assigned to their project who do not usually reside in the village with the students but make occasional visits. These evaluations include observations of projects, interviews with key village leaders, and review of a final report prepared by the students which describes the village, current problems in it, and students' activities and project outcomes.

One important component of service-learning programs according to the literature is "reflection" by participants on what they are learning from their giving of service. Neither of these programs specify a "reflective" component per se. However, there are many opportunities for the students in the Indonesian teams and the students and faculty in the Costa Rican projects to meet, talk, make decisions, and solve problems together. Thus, there is an emphasis on being thoughtful about what they are doing as they address real problems of their communities. But reflection is implied rather than highlighted. Also, the focus is on encouraging an orientation toward helping others rather than on self-conscious service (e.g., what am I getting out of this?).

Perceived Outcomes

To facilitate interpretation of the findings of these two studies, overall results as perceived by participants and by us as visitors are presented in Table 3 then discussed briefly.

The first substantial finding of this investigation was that the goals of both KKN and TCU are actually being translated into practices and experiences of faculty, students, and members of the community. It became clear that the service missions of these programs are being expressed in tangible, though somewhat different ways.

Table 3
Perceived Outcomes for the Indonesian and Costa Rican Programs

Area of Perceived Outcomes	Indonesia	Costa Rica
1. In terms of official objectives	The first three objectives are clearly being met. However, although relations among participating universities and communities are improving, it is not clear that the universities are adapting their teaching and research activities to the demands of the developing society.	All four objectives appear to be addressed by the current program. The service mission of TCU is being actualized into practices and experiences of faculty, students, and members of the community.
2. Relative emphasis is on community outcomes	The primary benefit is what the villagers receive from their involvement in KKN projects-- direct help with health, agricultural, educational, and other challenges as well as increased status for being a KKN project village.	While students, faculty, and university are benefiting, the explicit emphasis is on community benefits. The community members felt listened to and responded to by university representatives. They were invited to collaborate with students and faculty to address their own problems in small steps over time.
3. Student outcomes	In contrast with other coursework, KKN participation is perceived to develop self-management and interpersonal skills, and values, perspectives, and ideas related to social responsibility and multicultural awareness. It is also one rite of passage into adulthood.	With few exceptions, students feel that participation in TCU helps them develop civic responsibility and caring for others in society while they refine their skills and apply knowledge in their majors. They see this as one way to begin "paying back" their fellow citizens for this opportunity for a higher education.

4. Faculty/university outcomes	The universities which participate receive greater visibility in the communities involved and most graduates have a deeper knowledge of and commitment to the society; but faculty do not appear to benefit directly and curriculum and research programs are not affected directly.	Faculty who choose to participate report professional growth and learning while noting that participation helps them be better teachers and researchers and to integrate those responsibilities with service in meaningful ways.
5. Interdisciplinary focus	Students from many different disciplines are organized into teams which are assigned to address realistic project problems in villages.	The focus of TCU projects on real problems faced by communities leads naturally to integration of various disciplines to cooperatively address those problems.
6. Project focus	Use of original service projects to coalesce student and faculty effort in groups around community problems helps focus their efforts in ways that could not be achieved through solo service hours spent in existing service agencies.	Use of original service projects to coalesce student and faculty effort in groups around community problems helps focus their efforts in ways that could not be achieved through solo service hours spent in existing service agencies.
7. Placement in curriculum	Service in KKN as a requirement separate from all other coursework emphasizes the importance of the community needs in conjunction with university requirements.	Service in TCU as a requirement separate from all other coursework emphasizes the importance of the community needs over university requirements and makes the use of student time much more flexible.
8. Role of compulsion	The compulsory nature of the program is critical to its success but carries some problematic side effects.	The compulsory nature of the program is essential to its success but may lead to problems if other universities don't require it.

9. How well the program meets the Principles of Service Learning	Clearly meets 6 of 10 but does not emphasize critical reflection (Principle 2), often fails to clarify the responsibilities of participants (Principle 5), often lacks sufficient supervision of students (Principle 8), and provides insufficient time for many projects to be completed satisfactorily (Principle 9).	Clearly meets 8 of 10 but does not emphasize critical reflection (Principle 2) nor articulation of learning goals (Principle 3).
10. Main lesson learned	The focus on serving takes precedence over a focus on learning; this seems to be a strength rather than a weakness of the program.	The emphasis on social action for others by students and faculty leads to meaningful learning without (and maybe better than) an explicit focus on learning benefits to participants (particularly students).
	1. Faculty supervisor to student ratio is too small-- due to compulsory nature of the program. They need a mechanism to release faculty from other responsibilities in order to be placed with students in villages longer. 2. Lack of funds-- undue load carried by the students. 3. Length of program is too short for students to see consequences of their service. 4. Disagreement over the purposes and foci of KKN projects between students and villagers-- often an overemphasis on highly visible projects involving physical labor to	1. Quality of students' experience depends on the faculty member, the project, and the student and appears to vary considerably. 2. Compulsory element can lead to a negative experience for some students who would rather not participate. 3. There is a potential threat from other universities which do not require service; they may take over the higher education market. Therefore, there is growing pressure to justify TCU with respect to student benefits. 4. There is a need to involve more faculty in the program. While TCU provides an infrastructure for integrating research, teaching

11. Problems faced	satisfy villager's expectations while under emphasizing other student roles. 5. Institutionaliation of the KKN by the central government facilitates the amassing of resources and support from many sources but it can also lead to a focus on governmental goals and projects rather than local interests. 6. They need to develop more urban projects instead of working only on rural problems. 7. They need more external evaluation of their programs, including research on processes and outcomes of programs.	and service, how it is used depends largely on the academic departments' understanding of and commitment to the integration possibilities of TCU projects. Increasingly TCU staff are feeling the need to be able to demonstrate more convincingly the benefits to faculty and the university which emerge from TCU involvement so more departments will support research through this program and not require faculty to do research in competing areas. 5. Funding for projects needs to be increased. 6. They need more external evaluation of their programs, including research on processes and outcomes of programs.

Second, while students, faculty, and the university seem to be benefiting in a variety of ways, both programs are distinguished by their explicit emphasis (in terms of stated objectives and the planned activities of participants) on community benefits with more subtle or implicit emphasis on benefits to students, faculty, and the associated universities generally. Members of the communities being served by the university students and faculty reported that they felt understood and listened to by the academics. They were usually very involved in identifying their needs, developing the programs that would address those needs and evaluating their success in collaboration with the faculty and students involved. As one Costa Rican community leader said,

> This project has been good. . . it has helped people to get involved, to listen to one another. Some of our people have volunteered with building a canal here, because of this project. This land, you know, was given to us. This is a problem.. People sometimes begin to think other people should always solve their problems. This project has begun to show them that outside help is important, but it requires effort from us. Change requires collaboration. The best outside help is like this, when they help us decide, help us plan, help us accomplish our goals.

Third, however, with only a few exceptions, most of the university students felt that participation also helped them develop a realistic sense of civic responsibility, understanding, and caring for others in their society while providing an opportunity for them to refine their skills and apply knowledge obtained through their studies in new and useful ways. As one Indonesian student noted,

> I never realized that the villagers might have a set of customs and beliefs about what materials can be used for something like a water system and that concrete was not a part of that. This required a great deal of discussion. At times I was frustrated that we could not proceed faster. But I also gained an appreciation for their ability to deliberate on these matters. Without this sort of deliberation, important aspects of the culture could be lost. ... those ways might actually be there because they are the most effective. I suppose I did not realize that slow progress may sometimes be appropriate. ... KKN helped me learn the importance of examining these cultural issues, especially when drastic changes are being suggested. They know their reasons for previous practices and they know some of the barriers to change which outsiders can not anticipate."
>
> Another student from Costa Rica explained,
>
> I'm a boy from the city, but now my heart beats to the rhythm of the village. ...now I know the condition of many people in my country which I could have easily been shielded from my entire life. It is possible, you know, to pursue your own interests above the needs of people who are suffering nearby. We learn to navigate around them, those people, so well that we remain completely ignorant of their circumstances. But once you know about them first hand, ...then you no longer can only pursue your own interests. TCU helps you accept some responsibility for your society.

Fourth, Costa Rican faculty who choose to participate are also affected positively while Indonesian faculty are more tangential to their program. The universities in Indonesia which participate receive greater visibility in the communities involved and have the satisfaction of believing that most of their graduates have a deeper knowledge of and commitment to the society; but faculty do not appear to benefit directly and most curriculum and research programs are not affected directly by the service component.

In contrast, TCU appears to serve as a sort of faculty and university development program for at least a small proportion of the faculty. Participating faculty reported they benefited in many of the same ways students did, which was not surprising because one of the most obvious elements of faculty involvement across all Costa Rican sites was that faculty viewed themselves as learners alongside the students. As one professor reflected,

> This project gives me a chance to go to areas I never would otherwise go. I meet people I never would know and learn about circumstances I have only heard about. It is one thing to have an opinion or a solution in your mind. It is another to confront it, to address the problem, just as it exists, with all of the complications. ... And to do it in the peace of high mountain villages. Well, it is enjoyable and difficult at the same time. It changes you, personally and professionally. It anchors who you are and why you do what you do.

Others noted that participation helped them be better teachers and researchers and to integrate those responsibilities with service in meaningful ways:

> By blending them, I am not as torn between my responsibilities as I would be if I had

> them all separated.
>
> Doing different things is a good way to stay interesting to your students. Somehow I think I am more lively now when I teach my courses because TCU has gotten me out into the fresh air, off campus. TCU is a good reminder of what I am really doing as an instructor. It is easy to forget about that.
>
> For me it has been a good way to integrate research and service...and teaching. It has given me many ideas for teaching and research. I am very excited about this research. It is full of surprises.

One of the striking features of the research being conducted in the context of TCU projects is the way research questions arise. Many of the ideas for research projects emerge in the contexts in which they are being executed. Rather than formulating a research agenda out of context and imposing it on "subjects", the requirements of TCU call for projects which arise from community needs. For many researchers, this calls for a methodological departure often moving from highly reductionistic, quantitative approaches to more individualistic, qualitative research approaches. Several faculty find themselves developing new research skills as they ask different questions, often more directly bridging their research with practice.

Inasmuch as faculty experience personal or professional development, it is likely the university itself is indirectly being upgraded. Faculty indicated that the most apparent way the university benefits from their TCU experiences is that the curriculum is informed by the community-based projects. The strengthened community relationships which result from many TCU projects, including collaboration with various public and private agencies, also are perceived as beneficial to the university. And a student explained that as the faculty have improved through participation in this program, the whole university has benefited too,

> TCU makes the university more realistic. It forces faculty to take their theories into the streets, not just to test them, but to use them. And I think that when they teach, their experience of having used a certain theory or method in a real situation makes their lesson more meaningful to students, more valid. This improves the university generally, I think..

Fifth, programs in both countries address real problems of their national and local communities using interdisciplinary teams of students. In Indonesia, this occurs because the teams are created through a mixing of students from a variety of majors; then the teams are assigned to villages where the problems to be addressed are negotiated with the villagers and the students are able to call upon their experiences and coursework to address these real needs.

In Costa Rica, the faculties' focus in their TCU projects on real problems faced by communities throughout the country has lead naturally to the integration of various disciplines to cooperatively address those problems. The students join the projects from many different majors after reviewing brief summaries of the projects and in response to requests from the faculty members for students with particular expertise.

In both situations, the interdisciplinary efforts are breaking down barriers between people with different perspectives while identifying better solutions to problems they all face. As one student said,

> I am from the hard sciences. It was helpful to work with social workers on this. I would not have thought to do this, maybe because I was never taught to think that way, that inclusively, in my coursework. They helped me understand why people were making the decisions they were about how they built their houses; why they did or did not take

> care of their land like I thought they should. This was necessary to understand if we wanted to get them to change how they did things. It was humbling, I guess you would say, to realize that our solutions were useless until we learned how to reach them. This is why TCU is valuable to me. It is my work to help solve human's problems, but in my school, we do not study humans. We study the earth. TCU taught me the human side to geology that never appeared in my books.

And a faculty member noted,

> I have learned many things from this project already. I am a social worker, but I have had to learn about geology and geography, even architecture in order to make the social changes I am interested in making there. None of us can work in isolation when we are applying what we do to real life problems. This has helped me question my methodology and others' methodologies as well. What can it mean when people are working within a single perspective? What kinds of solutions to problems can they offer?

Sixth, in both countries, original service projects are created by the participants to coalesce efforts of all group members around community problems. This approach helps focus their efforts in ways that could not be achieved through solo service hours spent in existing service agencies. Students are not just giving service-- they are part of a team which supports them, challenges them, and helps them see that they are part of something bigger than themselves. They are learning to collaborate as citizens for the common good.

Seventh, although the Indonesian students are graded while the Costa Ricans receive pass/fail ratings only, both programs include the service curriculum as a separate requirement for students, outside of all coursework, again emphasizing the focus on community needs first and university structure and requirements second. Interestingly, several of the Costa Rican faculty who were interviewed as well as the TCU administrators, commented on how the aims of TCU depend on the freedom this external placement of the program provides. As one faculty member noted (and this seems to apply in Indonesia as well),

> I think it is better to have TCU separate as well as non-graded. It helps me achieve many of the objectives of TCU related to social responsibility. The interdisciplinary nature of TCU would be much more difficult if TCU were part of specific courses. And how can you respond to community needs if you have to do projects which must operate on the semester calendar? The way TCU works now, students are involved for 300 hours. How those hours are distributed depends totally on the nature of the work to be done for the community. It all depends on the community project. ...Having TCU non-graded is also very good, at least for me. It helped me use the evaluation process to achieve the objectives of social responsibility. If I had to grade the students, they would be motivated by a grade rather than by doing something which was meaningful to them and to others. The TCU process is just more real and so is the way they are evaluated. Don't you think it is good to learn how to evaluate your own work in your own terms? And, the students were more motivated than students ever are when grades are a part of it. This is ironic because some people, other faculty, who don't participate in TCU, say that it is impossible to motivate students if you are not grading them. Those of us who are committed to TCU do not find this to be true.

Eighth, one of the most hotly debated issues regarding the composition of university service-learning programs is whether service-learning should be cumpulsory or voluntary. Many of the elements in these programs which were cited as key in determining the value they have for their various participants are a function of their compulsory nature. The compulsory nature of the programs appears to be critical to their success; but this characteristic also bears some problematic side effects. In Indonesia, because nearly all universities participate in this national program, there are more students needing supervision and guidance than the existing faculty can appropriately serve. So, although many villages are receiving help, the quality may suffer.

In Costa Rica, in addition to the requirements placed on the students, each academic department at the University of Costa Rica is expected to have at least one TCU project underway at all times. Most participants seem to agree that the compulsory nature of the program has been essential to its overall success. A staff member who has been in the TCU program since its inception summarized,

> Once we had general commitment to TCU, we were able to be more creative in what we expected of students. When you require them to be in TCU, you can require them to be creative and to have the more intense experience, the better experience... maybe better than they would choose if you allowed them to choose it or not.

However, as mentioned earlier, as other universities which do not have this requirement continue to emerge and to compete for student enrollments, the compulsory nature of TCU is coming under increasing attack both from within and outside the university.

Ninth, an analysis of the Indonesian and Costa Rican program in terms of the ten principles of good practice for combining service and learning cited earlier (Kendall and Associates, 1990, p. 40) suggests both programs confirm and conform to most of these principles. It appears that eight of them were clearly represented in the experiences of participants in the TCU projects examined in this study. For example, relating to the seventh principle of service- learning, "An effective program expects genuine, active, and sustained organizational commitment," it appeared that TCU enjoys considerable institutional support. This support includes the compulsory nature of the program, the requirement of each academic department to be involved in TCU, financial and administrative support, and the clear articulation of TCU's mission and its relationship to the overall mission of the university.

However, in both countries there is evidence also that they do not follow these principles exactly. For instance, the Indonesian program clearly meets six of the ten but does not emphasize critical reflection (Principle 2), often incompletely clarifies the responsibilities of participants (Principle 5), lacks sufficient supervision of students (Principle 8), and provides insufficient time for many projects to be completed satisfactorily (Principle 9). Although the Indonesians would agree that they should meet the latter three principles better and they are searching for ways to do so, they do not appear to value the critical reflection component as much as the authors of the principles do.

Likewise, in Costa Rica, there is very little evidence that principle 2 (critical reflection) or the learning focus of principle 3 (articulation of clear learning goals) were explicitly addressed at all, almost in direct relation to the major emphasis of the program on "community" benefits. While there are many references in goal statements and program objectives to helping the communities associated with the university, neither of these principles that focus on more typical concerns of universities for their students are expressly manifest in the TCU program.

From many service-learning educators' perspectives, the absence of these student-centered components would be viewed as a programmatic weakness. In fact, some service-learning educators (see several in Jacoby, 1996) suggest that it is the presence of some of these components which distinguishes service-learning from non-educational volunteerism. They argue that one can volunteer but not necessarily interpret their volunteer experience accurately or ethically and not necessarily learn anything new without a reflective component and clearly stated learning goals. Thus, the absence of explicit student learning objectives in the TCU program and a reflective component in both programs may technically be noted as a weakness of these programs.

Paradoxically, however, it is possible that the lack of these particular ingredients may actually facilitate the objectives of both programs while better addressing the learning goals of students and faculty than the principles could, in this context. Perhaps to include components focusing on student learning objectives and a student- centered reflective process would be somewhat contradictory to the underlying philosophy of what the TCU and KKN experiences "ought" to mean to students since they emphasize an "other" orientation; and a focus on student outcomes emphasizes a "me" orientation. One of the TCU staff, pretending to be a student, tried to illustrate this point:

> TCU is not about me. Its about other people. Its about my community. ...Or the environment. It is not about ME!

The implication is that being a part of some action or activity which is other focused is a form of learning itself. Conversely, to demand the setting of student learning goals and objectives and the use of a formal reflective process for students to examine their experiences could be to undo the essence of the experience TCU and KKN are designed to facilitate. The exclusion of such elements may be as vital to the success of these programs as their inclusion in programs with different learning aims. The question which some service-learning educators will inevitably ask is whether the benefits students receive due to the absence of those elements can be called "learning".

When speaking about student experiences, TCU and KKN staff do not talk about learning as much as they talk about "social action" and service. In university settings where behavioral components of learning rarely appear, can such action constitute learning in and of itself? Some vehemently argue that the conscious processing of action is what constitutes the learning and is what may ensure that the actions have been "appropriately" interpreted and incorporated into the students' ethical and information systems.

Finally, in terms of challenges, participants in both countries are searching for ways to involve more faculty, to clarify the purposes of the program for potential participants, to obtain greater supporting funds, to improve the quality of the experience for all participants, to overcome negative side effects associated with their programs being compulsory, to have external evaluation assistance, and to conduct research on the processes and outcomes of their programs.

The Indonesians are also seeking to develop more urban projects rather than focus only on rural problems, and they are searching for ways to place faculty directly in field settings with students. The University of Costa Rica has found ways to deal with both these problems but finds that other newer universities throughout the country are not requiring students to give this kind of service and so they may lose enrollments over time if the program doesn't become a national one.

Lingering Questions

Our inquiries into the service learning programs of Indonesia and Costa Rica have clarified many of the issues that educators everywhere ought to consider as they contemplate similar programs in their contexts. But we have also encountered several questions that merit the attention and considered thinking of these educators as well. Some of these follow:

1. What are the relative advantages and disadvantages of attempting to develop or sustain a university program which is "action" oriented versus "learning" oriented?
2. What adaptations to the university mission are necessary to incorporate social action into the educational agenda?
3. What are the advantages and disadvantages (either in terms of learning or action) of

including elements in programs which encourage student-centered thought or reflection?

4. What types of educational processes are helpful and for what purposes?
5. How to create a balance between national and local community needs?
6. How to decide if enough student learning is occurring in a program, how to determine what learning is occurring, and how to expand learning benefits (if so desired).
7. What is the relationship between various programmatic elements and students' learning outcomes, including relative focus on the importance and function of training, monitoring, evaluation processes, reflective components, and concomitant traditional study?
8. How should teaching, research, and service be blended most appropriately.
9. How to demonstrate most convincingly the benefits to faculty and the university which emerge from involvement in service learning? What these benefits are, what elements are necessary to produce them, and how they can be documented and presented represent important research questions which need to be addressed.
10. What are the unexpected outcomes or side-effects of service learning programs?
11. To what extent are the experiences of the community, students, faculty, the university (i.e.. benefits, involvement, challenges) dependent on whether the program assumes a service project focus or consists of solo service activities? In what ways is the meaning of the service- learning experience for each participant influenced by this aspect of the program's design?
12. How do community, student, and faculty experiences differ depending on whether the service-learning activities are embedded in or external to traditional coursework?
13. What other aspects of participants' experiences are a result of the compulsory nature of programs and how do participants' experiences compare with the experiences of those involved in voluntary service-learning programs?

These questions mark important needs for future investigation.

Conclusions and Implications

Universities all over the world have continually aligned themselves with a three-pronged mission which includes research, teaching, and community service. While the ways in which each of these missions are defined may vary significantly from one university to the next, the priority service usually receives in comparison with the other two is quite consistent across many different university contexts in which service at least ostensibly appears as a part of the mission. Almost without exception, service in academic life remains a low priority, is often ambiguously and narrowly defined, frequently refers only to on-campus service within or between departments, rarely is integrated into research and teaching activities and involvement with students, and is rarely considered as an equally viable component in promotion and tenure evaluations.

The KKN and TCU programs serve as examples of university programs which promote service, though still as a "third" priority, as a more legitimate part of students' and faculties' academic lives than in most universities in the world.

Increasingly universities everywhere are receiving the message that students need to be developing a greater sense of community membership, interdisciplinary understanding of social problems, and an enhanced ability to apply their knowledge and skills to a wide variety of circumstances. Communities which host universities often feel isolated from the university members, frequently serving only as subjects for their research, and often removed from the wealth of resources housed in those "ivory" towers. As a result, universities are increasingly re-evaluating their service missions and asking the question, "Why is service part of the university's mission and what does it mean?"

In this study, we have begun to examine the meaning of the KKN and TCU experiences as well as challenges they face from many different participants' perspectives. While the focus of both programs seems to be on benefits to the community, participants also acknowledged that better linkages need to be forged between students' learning needs, faculty needs, and the needs of communities in which students and faculty are involved if more faculty are going to get involved and if financial support from various sources is going to be forthcoming.

It seems fair to conclude from our experiences in these two countries that the key variable to overcoming challenges and expanding service-learning beyond the levels that have been achieved is increased faculty involvement. The communities are already benefiting and welcoming whatever the universities are offering. The students participate because it is mandated. But until more faculty are involved by helping them understand both the "why" and the "how" of service-learning, both these exemplary programs are limited. The Indonesian program is unable to grow due to lack of supervision while the Costa Rican program remains fairly small because fewer than 20% of the faculty are involved in developing projects in which students may engage.

In an effort to understand the faculty role better, we asked faculty who were involved why they considered involvement a legitimate expenditure of their time and energy. However, their answers to this question usually did not quench our curiosities. Their responses were,

> "We must."
> "It is important, don't you think so?"
> "It is interesting."

It began to be clear that while the support mechanisms associated with these programs were necessary and helpful, they were not what made the prospect of involvement compelling for faculty. All faculty in the University of Costa Rica and the Indonesian system are repeatedly provided with this information, yet only about one fifth of the faculty participate. Why do some choose to get involved and view participation as an ideal way to combine teaching, research and service while others simply don't participate?

We were reminded of the philosophy of Emmanuel Levinas (1987) who suggests that when one truly comes to acknowledge another person and their call upon the acknowledger, engaging with that "other" ethically comes naturally, as do methods for responding to the call of the other. This view prompts us to wonder if in our study of methods used in these two service-learning programs, perhaps we only brushed the surface of the most important characteristics of KKN and TCU: the reasons why some faculty (and universities) are committed to these programs; and, conversely, why some faculty, even though they have the same information about how these programs operate, choose not to be involved?

Is it possible that the critical issue is one of philosophy or paradigm; how those involved with service- learning see the world differently than those who are not involved? Perhaps what is needed to expand these programs is vision--the assumptions of individuals and their organizations in relation to the communities which surround and support them. The question becomes, "How can this vision be shared?"

Some faculty criticize these programs as free labor given to the community. Reflecting on this perceived problem, a TCU staff member stated,

> We need ideas for educating our faculty, for motivating them about TCU in a positive way. Clearly they do not understand TCU when they feel that way. TCU is not free labor. We need a way to show them how educational it is for students. But its more than that. I don't know. It is difficult to explain. But we need this more and more because their sentiment seems to be growing."

They need more of what? As qualitative researchers who have spent considerable time exploring the paradigmatic assumptions of this form of inquiry, we recognize significant parallels between the assumptions of a qualitative inquiry paradigm and the underlying assumptions of service-learning. For example, qualitative inquirers assume that to develop an understanding of another phenomenon such as a person, it is necessary to interact with them from multiple perspectives, to allow the phenomenon to affect oneself. Qualitative researchers also assume there are no simple causal relationships but rather complex interrelationships that generally require interdisciplinary perspectives to be understood, that values influence all constructions of knowledge (simple objectivity is impossible), that constructs about phenomena should be defined within their own context and language, and that to influence phenomena with any amount of deliberation, all of these assumptions must be encountered.

We also noticed, as we have in our work with researchers (whether quantitative or qualitative), a sort of blindness on the part of "insiders" in these service-learning programs regarding the assumptions they were operating under. Similar to researchers we have worked with, the "insiders" were often unable to articulate their assumptions clearly. And so, it is perhaps this inability to understand one's own assumptions that handicaps them in being able to expand their programs to other faculty and associated students and community members. For example, a person who has not considered their teaching, learning, and research paradigms, especially if they contrast drastically with those of participants in service-learning programs, would likely find statements like, "TCU is a must" or "KKN is important" not very convincing arguments for engaging in these programs. But it is those kinds of statements faculty tend to make regarding the impetus for their involvement.

If what we have found in our experience in helping researchers consider alternative paradigms of inquiry relates to the process of helping faculty consider their entire professional paradigms, it may require an in-depth examination of their paradigmatic assumptions in order for them to understand what service-learning is based on and why they should be involved.

A subsequent related question is, what risks are involved in bringing these assumptions into the primary awareness of those persons already committed to and engaged in service-learning? Polanyi (1962), in *Personal Knowledge* discusses the value of both primary and subsidiary awareness and what elements of experience belong to each. He offers the example of a pianist who suddenly pays attention to his fingers and immediately falters. While sometimes it is helpful to focus on the fingers, in developing form for example, at other times it is not. Addressing the point that some "information" belongs only to the realm of tacit knowledge, Polanyi also points out how explicit mathematical descriptions of what is involved in keeping a bicycle balanced as it is ridden down a sidewalk do not help the bicyclist accomplish this task.

Therefore, pulling up the underlying assumptions of what fuels commitment and understanding of service-learning programs such as TCU and KKN, may create self-conscious servers for the future and may not actually lead to others' participation. But, our experience in helping researchers consider alternative paradigms of inquiry suggests that such examination can be extremely liberating for people and can help them make choices they never consciously made before. Furthermore, by understanding those assumptions, they are less likely to violate them and, consequently, are more likely to be successful within the parameters upon which the paradigm enables them to act.

TCU and KKN represent rich fields of inquiry in which the principles of good practice for combining service and learning can be explored. This inquiry has demonstrated the need for research in many different areas regarding university service-learning. This inquiry of TCU and KKN also serves as evidence that there is great potential value in looking in unexpected places for innovation and the need for educators to look to educators in developing countries to see what they can teach us for a change.

References

Albert, G. (Ed.) (1994). Service-learning reader: reflections and perspectives on service. Raleigh, N. C.: National Society for Experiential Education.

Arias Sanchez, O. (1988). We seek ... Not peace alone. Commencement address at Harvard University, *Harvard Magazine*, July-August, 19-22.

Daloz, L. A. P., Keen, C. H., Keen, J. P., & Parks, S. D. (1996). Common fire: lives of commitment in a complex world. Boston, MA: Deacon Press.

Denzin, N. K. & Lincoln, Y. S. (1994). *Handbook of qualitative research.* Thousand Oaks, CA: SAGE.

Direktorate Pembinaan dan Pengabdian pada Maysarakat Ditjen Kikti Depdikbud, 1986.

Eberly, D. J. & Sherraden, M. (Editors) (1990). *The moral equivalent of war? A study of non-military service in nine nations.* New York: Greenwood Press.

Eisner, E. (1991). *The enlightened eye: qualitative inquiry and the enhancement of educational practices.* New York: Macmillan.

Gonzalez, M. A. (1992). In Eberly, D. J. (Ed.), *National youth service: A global perspective.* Washington, DC.: National Service Secretariat.

Hardjasoemantri, K. (1981). Study-service as a subsystem in Indonesian higher education. Unpublished dissertation.

Información General (1992). (General Information). San José, Costa Rica: The University of Costa Rica, Vice Rectory of Social Action.

Jacoby, B. and Associates (1996). *Service-learning in higher education: concepts and practices.* San Francisco, CA: Jossey-Bass Publishers.

Kendall, J. C. & Associates (Eds.) (1990). Combining service and learning: a resource book for community and public service, Volume I. Raleigh, NC: National Society for Internships and Experiential Education.

Polanyi, M. (1962). *Personal Knowledge: Towards a Post-Critical Philosophy.* Chicago, IL: University of Chicago Press.

Sherraden, M. & Castillo, C. M. (1990). "Costa Rica: Non- military service in a nation with no army." In Eberly, D. J. and Sherraden, M. (Editors), *The moral equivalent of war? A study of non-military service in nine nations.* New York: Greenwood Press.

Acknowledgment

This project was supported by Fulbright for visits to Indonesia and by the National Service Secretariat for visits to Costa Rica.

Note

All quotes are based on actual comments made by interviewees; however, because most of these people spoke Spanish, they are not literal quotes. Rather, they have been embellished by us to reflect the spirit of what they said.

About the Authors

David D. Williams

Brigham Young University

e-mail: David_Williams@byu.edu

David D. Williams is an Associate Professor in the David O. McKay School of Education at Brigham Young University. He earned his PhD in Educational Research and Evaluation Methods at the University of Colorado, Boulder in 1981. His M.A. (1978) in Educational Leadership (emphasis in evaluation) is from Western Michigan University; his Bachelors degree (1976) is from Brigham Young University.

Dr. Williams holds an appointment in the Department of Instructional Psychology and Technology at BYU, where he teaches graduate courses in qualitative inquiry, research methods, and educational evaluation. He conducts research in the areas of service and experiential learning, international education and development, integrated curricula, qualitative methods of research and evaluation, use of computers in education, and the role of the learner in the education process. He is currently conducting an evaluation of an innovative teacher preparation and professional development program in Utah.

See David Williams's Homepage for more details. Visit his qualitative inquiry class homepage.

William D. Eiserman

Colorado Foundation for Families and Children

Jivio@aol.com

William D. Eiserman earned his Ph.D. in Instructional Design at Brigham Young University in 1986. Currently Dr. Eiserman is on the research faculty in the Department of Communication Disorders and Speech Sciences at the University of Colorado-Boulder. He also has a joint appointment at the Colorado Foundation for Families and Children in Denver, Colorado. Dr. Eiserman's research focuses on the needs of families as they adapt to the birth of a child with a disability and on the various ways communities support these families and children. He also has a strong interest in service and experiential learning and has engaged in research on this topic in a variety of international contexts, one of which was initiated while on a Fulbright Fellowship in Indonesia.

Copyright 1997 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://olam.ed.asu.edu/epaa

General questions about appropriateness of topics or particular articles may be addressed to the

Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey@olam.ed.asu.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

John Covaleskie
Northern Michigan University

Andrew Coulson
a_coulson@msn.com

Alan Davis
University of Colorado, Denver

Sherman Dorn
University of South Florida

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
hmwkhelp@scott.net

Thomas F. Green
Syracuse University

Alison I. Griffith
York University

Arlen Gullickson
Western Michigan University

Ernest R. House
University of Colorado

Aimee Howley
Marshall University

Craig B. Howley
Appalachia Educational Laboratory

William Hunter
University of Calgary

Richard M. Jaeger
University of North Carolina--Greensboro

Daniel Kallós
Umeå University

Benjamin Levin
University of Manitoba

Thomas Mauhs-Pugh
Rocky Mountain College

Dewayne Matthews
Western Interstate Commission for Higher Education

William McInerney
Purdue University

Mary P. McKeown
Arizona Board of Regents

Les McLean
University of Toronto

Susan Bobbitt Nolen
University of Washington

Anne L. Pemberton
apembert@pen.k12.va.us

Hugh G. Petrie
SUNY Buffalo

Richard C. Richardson
Arizona State University

Anthony G. Rud Jr.
Purdue University

Dennis Sayers
University of California at Davis

Jay D. Scribner
University of Texas at Austin

Michael Scriven
scriven@aol.com

Robert E. Stake
University of Illinois--UC

Robert Stonehill
U.S. Department of Education

Robert T. Stout
Arizona State University

---

archives | abstracts | editors | board | submit | comment | subscribe | search

EPAA • enter the archives • browse the abstracts • the editors • the edit board
• submit article • submit commentary • search • subscribe
volume: • 1 • 2 • 3 • 4 • 5 • 6 • 7

This article has been retrieved 1269 times since February 9, 1997.

Education Policy Analysis Archives

Volume 5 Number 5 February 9, 1997 ISSN 1068-2341

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass Glass@ASU.EDU.

College of Education

Arizona State University,Tempe AZ 85287-2411

Copyright 1997, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

Out of Our Minds: A Review

Pat Clifford
Banded Peak School
Bragg Creek, Alberta

Howley, Craig B., Howley, Aimee & Pendarvis, Edwina D. (1995). *Out of Our Minds: Anti-Intellectualism and Talent Development in American Schooling* NY, NY: Teachers College Press. Pp. 265; $24.95 (Paper)

Out of Our Minds: Anti-Intellectualism and Talent Development in American Schooling will probably make a lot of people mad. Conservatives who pick up the book on the basis of title alone, expecting a rant-and-canon, will hate it. Radicals attracted by the authors' exploration of the role of capitalism and its imperatives in the intellectual impoverishment of students and teachers may find themselves discomfited to find that critical theory becomes, in places, the subject of the book's critique. The authors, Craig and Aimee Howley and Edwina Pendarvis, have already been accused of teacher-bashing in reviews of early drafts of Chapter 2, which exposes the ways in which teachers and schooling (including the university) are key mechanisms of anti-intellectualism, not its remedy. Part of a larger Teachers College Press series devoted to Education and Psychology of the Gifted, the book is bound to raise the hackles of those whom James Borland, the Series editor, calls "steadfast defenders of the faith who regard any challenge to 'gifted programs' as typically constituted in educational practice as errant and pernicious nonsense" (vii).

For others, however, that delicious set of disappointments may well constitute the very attraction of this book. There is no shortage, these days, of exposés of the failure of schools. As a teacher and researcher, I am generally attracted to these books the way I am attracted to smash-ups on the freeway: I don't want to look, but I can't seem to help myself. All too often, critics and their remedies follow either predictable ideological grooves or hawk warmed-over fads from schools of management *Out of Our Minds* isn't like that, however. From the outset, its tone is both honest and impassioned. Admitting, in the Preface, that the book captures both the anger and the arrogance that permits them to write about matters of anti-intellectualism in schools at all, the authors touch on a sore point for grouchy, between-the-fingers readers of exhortatory texts on school improvement: How dare someone sitting in a comfy office miles away from the mess and complication of my life as a teacher write about what I need to do to improve myself? How dare they?

The authors get that out of the way right from the start. There is an arrogance in presuming that what you believe, is good for another to know. No sense scuffing the toe of your academic shoe in the dust and pretending it ain't so. As a teacher who also writes, I had to admit my own necessary arrogance in sending words about my work in classrooms out into the world: How dare I write these neat and tidy pieces about the wonderful work my students do (okay, okay, the wonderful work I do as their teacher) as if there were not moments, hours, sometimes days and weeks at a stretch of my own failure, stupidity and confusion. How dare I?

On the second page of the Preface, I was hooked. There is a presumption in writing about the work of others that has a shadow few will claim. Yet rage at the injustice that debases the lives of children in school makes such speaking out morally imperative at the risk of colonizing the lives of others; at the risk of being dismissed as teacher bashers; and at the risk (or so I find it in my own work) of exposing all the places in which one's own practice falls short and disappoints; at the risk of papering over exactly these holes and fissures. Claiming the territory into which they are about to move, the authors insist:

> that anger and arrogance ought not to disguise who we are in ordinary life: We have regular jobs, we have the usual troubles, we are too seldom able to do the right thing in terms that approach what we truly believe. We bite our tongues too often, despite the sharpness of the critique in this book. The dilemma we are writing about is our own.

And I like that. The stated aim of the book is not to sneak peeks at johnny-come-lately schemes for school improvement. Instead, its authors intend to invoke and to invite sustained conversation about educational reform that is emancipatory and deeply respectful of human intellect and potential, however conflicted and difficult the actual life of obligation that a call to such reform demands.

Structure of the Argument

The argument of *Out of Our Minds* is wide-ranging and complex, but its organizing motif is the refusal of schools to claim an intellectual mission. Placing responsibility for the institutional debasement of the life of the mind squarely in the political realm, Howley, Howley and Pendarvis point directly to "the things to which the economy, the polity, and the society ascribe value; in short, to the organization of production, governance, and social relations..." (xiii). The capitalist goal of global domination that has been so explicitly a part of American school reform efforts since Sputnik turns schools, they argue, into institutions whose central purpose is instrumental, not intellectual: the production of patriotic job-holders. They quote Michael Apple on "the threat that this goal for public education poses for the commonweal" (xiv):

> This transformation [of the purposes of education...]involves a major shift- -one that Dewey would have shuddered at--in which democracy becomes an economic, not a

political, concept and where the idea of the public good withers at its very roots.

And it is the public good that concerns the authors throughout. They point to the failure of schools to nurture the talents of the intellectually gifted as a particular educational travesty, arguing systemic neglect of human potential even in programs for the gifted, surely the place one might expect to find it most carefully guarded. However, their concern for gifted students is located in a larger concern for the intellectual fate of all children in American schools, especially those who, lacking birthright privileges of race, class, ethnicity and gender, are most denied access to social goods and to an education that values the life of the mind.

They argue, against the grain of much educational theorizing, that intelligence read as inborn ability, is not the salient construct, even in an analysis of the experience of gifted children in school. Regarding IQ as an artifact of educational psychology, they are not especially interested in quantifiable, instrumental and individualistic measures of practical performance designed to predict academic success. Rather, they insist that schools ought to be about the stewardship of intellect: the fostering of intellectual habits of thought, meaning-making and discourse in all students regardless of their putative (and arguably) inborn potential. Intellect, in this light, "represents the complexity of understanding, critique, and imagination of which the human mind is capable....Intellect [has] to do with what passes between minds and generations of humans...[through] explicit, negotiated meaningfulness" (p. 4).

For the authors, what passes between the minds and generations of students and teachers bound together in schools is thin gruel, indeed. Whether the schooling offered most children in America derives from traditional, conservative assumptions or more progressive, liberal ideology, the end result, the authors insist, is strikingly similar: an instrumental approach to the purpose and structure of schools that subjugates intellect to the service of practical action and the acquisition of the "little literacies" of problem solving and skill accumulation. A true education , they argue, has three characteristics glaringly absent from contemporary schools. respect for the interest of the intellect (contemplation, interpretation, understanding, meaning- making, and critique); respect for the artifacts of intellect (particularly the ability to speak and write well); and respect for the intellectual potential of all students.

Examining the origins of the anti-intellectualism that makes sustained dialogue between individual students and the historical community of learning almost impossible, the authors focus directly on the anti-intellectualism of teachers and the institutions of public education and the university. It is this analysis that early reviewers dismissed, incorrectly, as "teacher-bashing". As a teacher for more than a quarter century, I recognized the world they describe. It is a world in which few of my colleagues read much (they cite studies showing as few as three books a year, and those most often popular fiction), in which few make a special effort to pursue advanced study in a subject discipline, and in which the penalties for presenting oneself as an intellectual are swift and severe. It is a world in which the outspoken and unconventional (both child and teacher) invoke the ire of the organization. And this world of schools, they argue, is in cahoots with parents, the anti-intellectual university and society at large to ensure that future employment, not the cultivation of the life of the mind, is the major focus of education. It is a world in which bright students deal with schools about as well as other children do: that is, badly. And it is a world in which the pernicious effects of racism, sexism and poverty stunt the development of intellectual potential.

Throughout, the authors take steady aim at important targets: the elitist notion that neglecting gifted education means neglecting society's brightest and best resources; the claptrap that has arisen around much school and curriculum reform literature over the past thirty years; and the insidiousness of research, theory and practice that support meritocratic explanations of career and academic success, particularly as such theory and practice ignores systemic discrimination against girls and women, visible minorities, and the poor. No one is spared, not even the remedies suggested critical theorists with whom the authors claim closest kinship in their analytical framework.

Rethinking the Potential of Schooling

I like the angry, overtly political tone of this book. My teaching and research partner, Sharon Friesen, and I have been in hot water for things we have written over the years, for intemperate outbursts at dinner parties, for stony silences in staff meetings around whatever reforms du jour are currently circulating through our school districts like snake oil. We get mad a lot. And somehow, stubbornly and intractably, we keep teaching anyhow, determined to make things different for the children in our care, having long since given up hope that systemic reform is actually possible, but called (naively, perhaps) to a vision of teaching and learning that makes a difference in the lives of children and their parents.

All through the reading of the early chapters, I disciplined myself from flipping to the part that I really wanted to read: the final chapters on suggested changes to schools and teaching practice. As depressing as I found the experience of revisiting my own frustrations with school through the authors' analytic framework, I resisted the temptation to rush ahead to the end. I'm glad I did, for the urgency of their demand for systemic, fundamental change provides a necessary background for understanding why they also believe (as do I) that the suggestions they offer as alternatives are unlikely to be implemented on any large scale. We are stuck with schools as we know them, they argue, either because we can't seem to imagine that they might, in fundamental ways, be otherwise, or because we cannot tolerate "the mismatch between citizens who are schooled in the intellectual habits of inquiry and critique and a political economy that is dependent on mute and compliant workers" (181).

And yet, as they point out, from time to time, in small pockets and in the practices of individual teachers who have found ways to resist and to create something new, we catch glimpses of how life in schools might be otherwise. The final chapters of the book tease out some of those possibilities in terms of their central image of stewardship. Invoking Norbert Weiner"s injunction in *The Human Use of Human Beings* (1950), the authors insist that it is time that educators get busy and figure out what human beings are for. Schemes, they say, that "try to warrant educational practices on the basis of trite claims about "future needs" and posturing for global domination are unethical, undemocratic, and ugly" (p. 184). Human beings are for making meaning, they insist, and the sacrifice of meaningful engagement with work, with one another and with nature that our narrowly instrumental institutions demand are, to these authors, nothing short of evil.

Stewardship in education, they say, "comprehends the need for humans to take pleasure in their work and to care for the human artifact. Stewardship also comprehends care for generations past and generations to come. This sort of stewardship is a commitment due students from teachers, children from parents, and the world at large from the people within it" (p. 185). Correctly, I believe, they insist that the ethical, political, and aesthetic choices people make determine what their world will be. Education is all about the quality of those choices.

Admittedly sketchy in their remarks about what classrooms devoted to the stewardship of intellect might actually look like, they nevertheless invite dialogue by sharing some of the principles they have glimpsed as they have worked with teachers whose practice gives them hope. They identify ways of thinking about "the problem of the canon": how do you decide which works are worthy of sustained engagement? How do you construct educational processes that honor "the fact that human beings are at work on themselves (p. 188, emphasis in the original)? Educational processes worthy of the name provide, they say, conditions for insight, vision, and ephiphanies that lead students not, as is now the case, to learn woefully less than we might hope, but to learn, instead, far more. And what students might learn in morally, politically and aesthetically reformed schools would come closer, they suggest, to the kind of schooling the world actually needs.

I found their ideas appealing and exciting because the principles they outline are ones my teaching partner and I have struggled to bring into being in our classroom for the past six years. We have begun to document our struggles and our successes, exploring and explaining what children's work looks like when we take seriously the search for answers to questions like the ones this book raises. Some of that writing is published in print, the most readily accessible being our article, "A curious plan: Managing on the twelfth" originally published in *Harvard Educational Review* in the fall of 1993 and most recently reprinted in Class Acts: Teachers Reflect on Their Own Teaching Practice.

For ease of reference in this review, I have also put three of our other papers on line: Hard Fun: Teaching and Learning for the 21st Century, originally written for the Government of Canada as part of a series of articles on innovative social policy; Landscapes of Loss: On the Original Difficulty of Reading a chapter in a forthcoming book on action research, and The Transgressive Energy of Mythic Wives and Wilful Children: Old Stories for New Times , first presented at the 1995 JCT Conference on Curriculum Theory and Classroom Practice and now under consideration for inclusion in a book on education and mythology. And last week, we put examples of Grade Two children's response to an extended study of The Odyssey on our school website.

I offer these examples in the spirit of teacher scholarship that Howley, Howley and Pendarvis describe. Schooling as they envision it "ought to cultivate a true education, assisting students to make personal and collective sense out of their encounters with the culture in both its intellectual and popular forms" (p. 190). This aim, they acknowledge, calls forth very different kind of teacher-scholars, ones who honor "through their customary practice a personal commitment to the life of the mind" (p. 191): who pursue scholarly pursuits in their own fields, and who take pedagogical encounters as opportunities for reflection and for action. At the beginning of the book, they express the hope that their readers will take part in "a conversation begun long ago and that must continue well into the future" (xii). They hope that others will be drawn into the conversation as they think, speak and act; as, indeed, I felt drawn in as I saw the possibility of the work Sharon and I do contributing a particular kind of flesh to the backbone of this book's argument.

And now, as I hope the readers of this review might themselves feel drawn. This on-line journal seems particularly well positioned to encourage actual, not just virtual, dialogue among educators; dialogue that welcomes the voices of children and teachers so conspicuously absent from much of the theorizing done on their behalf.

I think the authors of *Out of Our Minds* would approve.

About the Authors

Patricia Clifford

Banded Peak School
Bragg Creek, Alberta

http://www.rockyview.ab.ca/bpeak

Copyright 1997 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is
http://olam.ed.asu.edu/epaa

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey@olam.ed.asu.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

John Covaleskie
Northern Michigan University

Andrew Coulson
a_coulson@msn.com

Alan Davis
University of Colorado, Denver

Sherman Dorn
University of South Florida

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
hmwkhelp@scott.net

Thomas F. Green
Syracuse University

Alison I. Griffith
York University

Arlen Gullickson
Western Michigan University

Ernest R. House
University of Colorado

Aimee Howley
Marshall University

Craig B. Howley
Appalachia Educational Laboratory

William Hunter
University of Calgary

Richard M. Jaeger
University of North Carolina--Greensboro

Daniel Kallós
Umeå University

Benjamin Levin
University of Manitoba

Thomas Mauhs-Pugh
Rocky Mountain College

Dewayne Matthews
Western Interstate Commission for Higher Education

William McInerney
Purdue University

Mary P. McKeown
Arizona Board of Regents

Les McLean
University of Toronto

Susan Bobbitt Nolen
University of Washington

Anne L. Pemberton
apembert@pen.k12.va.us

Hugh G. Petrie
SUNY Buffalo

Richard C. Richardson
Arizona State University

Anthony G. Rud Jr.
Purdue University

Dennis Sayers
University of California at Davis

Jay D. Scribner
University of Texas at Austin

Michael Scriven
scriven@aol.com

Robert E. Stake
University of Illinois--UC

Robert Stonehill
U.S. Department of Education

Robert T. Stout
Arizona State University

---

archives | abstracts | editors | board | submit | comment | subscribe | search

This article has been retrieved 4813 times since February 12, 1997.

Education Policy Analysis Archives

Volume 5 Number 6 February 12, 1997 ISSN 1068-2341

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass Glass@ASU.EDU.

College of Education

Arizona State University,Tempe AZ 85287-2411

Copyright 1997, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

Academic Freedom, Tenure, and Student Evaluation of Faculty: Galloping Polls In The 21st Century

Robert E. Haskell

University of New England[1]

Abstract: Despite a history of conflicting research on the reliability and validity of student evaluation of faculty (SEF) it has typically not been viewed as an infringement on academic freedom. When it is suggested that SEF may impinge on academic freedom, it is often considered an attack on either student rights, or on the process of evaluating faculty performance in general. Faculty and educational administrator views and surveys are reviewed as SEF is used in salary, promotion and tenure decisions. It is suggested that the literature shows that SEF infringe on instructional responsibilities of faculty by providing a control mechanism over curricular, course content, grading, and teaching methodology. It is further suggested that SEF play a significant role in current attacks on tenure, and that its role in a demographically diverse 21st century educational system has changed from its benign historical origins. It is concluded that contrary to current views, SEF is a serious unrecognized infringement on academic freedom.

Despite the long line of research on the validity of SEF it has not typically been viewed as impinging upon academic freedom. At first glance, the very notion seems counterintuitive, if not downright mean spirited. Indeed, when it is suggested, it is seen as not only a novel idea, but as an attack on either students, or a general attack on evaluating faculty. In fact, one recent handbook for college administrators, (Weeks, 1996) refers to such an idea as "a rather novel attack on the use of student evaluations in assessing a faculty member's performance." Indeed, it has generally been taken for granted that SEF is appropriate and necessary. Unlike the body of research on the methodology of SEF construction and the validity of the ratings, there is a paucity of data on the issue of its impact on academic freedom.

The likely reasons for this paucity of data are that (1) there has been little professional mention of SEF as an infringement on academic freedom, (2) some faculty are embarrassed to admit that student evaluations may influence their professional behavior in the classroom, (3), to question the right of students to evaluate faculty may be considered unprofessional if not undemocratic, (4) to question the right of students to evaluate faculty may be seen as self serving, and (5)SEF tends not to be high status research. Finally when SEF is recognized to have an impact, unlike traditional threats to tenure and academic freedom, the deceptive appearance of SEF does not seem to warrant serious concern.

In a recent and otherwise carefully reasoned book on academic freedom (Menand, 1996), the issue of SEF is not mentioned.[2] At best, the subject of SEF has been the orphan in discussions on academic freedom with no direct logical lineage. At worse, it has been considered illegitimate. There are signs, however, indicating wider faculty recognition of this issue. As administrative policy, the use of SEF has largely evolved in a de facto manner, this paper will review the issue in search of a more tutored policy.

It is important to note at the outset, that it is not SEF per se that is the issue, but the impact of its use on salary, promotion, tenure decisions, and equally important, its impact on the delivery of quality education. In the evolution of any policy the accumulation of data, judgements, and arguments around an issue need to be coalesced. As the history of legal rights demonstrates, issues not considered to have legitimate standing only come to have standing after a long process of advocacy. It is, therefore, the purpose of this paper to coalesce arguments, and data, and explore implications of SEF scattered throughout the literature and to thus facilitate this developmental process.

In explicating SEF, many closely related issues must be substantially bracketed. These related issues include (1) its validity (Cahn, 1987; Damron, 1996; Greenwald, 1996; Greenwald and Gillmore, 1996; Scriven, 1993; Seldin, 1984; Tagomori, Bishop, and Laurence. 1995),[3] (2) the problem of defining teaching effectiveness,[4] (3) general variables affecting SEF scores, [5] (4) alternatives to SEF's (5) classroom politically correct or popular standards and perceptions, (6) low student academic preparation, (7) age and gender discrimination issues (Feldman, 1983, 1993) (8) strategies for change, and (9) other integrally related issues such as their being largely responsible for lowered course standards, and grade inflation.[6] Though these are important and related issues, they can only be addressed here in so far as they directly impact the focus on SEF and academic freedom.

While SEF was apparently first used in the early 1920's at the University of Washington, beginning in about the 1960's SEF has been increasingly used by universities in decisions on tenure and promotion. One of the reasons that SEF was instituted---and rightly so---was for informational feedback so that faculty might be more aware of student needs. The instrument has not, however, been used just for informational feedback to professors. If this were the case, then SEF would presumably not be a problem.[7] As Cashin (1996), Director of the Kansas State University, Center For Faculty Evaluation and Development, notes, "The higher education rhetoric is almost universal in stating that the primary purpose of faculty evaluation is to help faculty improve their performance. However, an examination of the systems--as used--indicates that the primary purpose is almost always to make personnel decisions. That is, to make decisions for retention, promotion, tenure, and salary increases." Herein lies the problem.[8]

In comparing two studies of the same 600 liberal-arts colleges, the author found that the number of institutions using student ratings to evaluate instructors had escalated from 29 per cent to 68 per cent to 86 per cent. The author noted that no other method of evaluation has approached that level of usage (Seldin, 1993). Another survey found that most business schools now use SEF for decision making, with 95% of the deans at 220 accredited undergraduate schools always making use of them as a source of information (Crumbley, 1995). Two nationwide studies of accounting department Chairpersons, indicated that reliance upon SEF was second only to research publications in professional journals (Yunker and Sterner, 1988). Department chairs and

Deans often weigh student ratings heavily in the faculty evaluation process. Perhaps no other method of evaluation has become so sacrosanct. SEF are used not only in the U.S. but in Australian, Canada, Europe and Great Britain. Unlike in the U.S., however, in Great Britain SEF by formal questionnaire, despite apparently no formal mandate, are increasingly used, though not weighed as heavily as is information gathered by other means. (Husbands and Fosh, 1993).

While considerable research---both past and present--- has been conducted to assess the validity of SEF, formal reports or studies suggesting that they might impinge on academic freedom are virtually non existent. There is likewise a paucity of direct hard data to support widespread but informal evidence addressing the issue of SEF as impinging on academic freedom.[9] Informal and reasoned analyses of the issue indicate that because SEF is used for faculty salary, promotion, and tenure decisions, there is pressure to comply with student classroom demands regarding teaching style, grading and a host of others demands(see below). It is suggested that it is this pressure to comply with student demands that directly leads to an infringement upon academic freedom.

SEF is not simply a salary, promotion, and tenure issue--- as important as these are for individual faculty; nor is the issue simply that students evaluate faculty. As the findings of this paper suggest, SEF are not the benign instrument they may appear to be or may once have been. Their primary impact goesto the core of academic freedom and to quality of instruction.

A Brief Look At Academic Freedom

Academic freedom and tenure are two sides of the same coin. Following an initial statement of principles in 1915, the current view of tenure was established in 1940 when the American Association of University Professors (AAUP) and the Association of American Colleges (AAC) officially sanctioned it for purposes of preserving faculty's right to academic freedom. Legally, it assures faculty the right to pursue any line of inquiry in the course of their teaching or research without being censored, penalized or fired by university administrators.[10] In 1973, the Commission on Academic Tenure in Higher Education (jointly sponsored by the AAUP and the AAC) made the following recommendation:

> The commission believes that "adequate cause" in faculty dismissal proceedings should be restricted to (a) demonstrated incompetence and dishonesty in teaching and research, (b) substantial and manifest neglect of duty, and (c) personal conduct which substantially impairs the individual's fulfillment of institutional responsibilities. The burden of proof in establishing cause for dismissal rests upon the institution (University of Michigan, 1994).

The AAUP's Statement on Teaching Evaluation suggests that

> Casual procedures, a paucity of data, and unilateral judgments by department chairs and deans too often characterize the evaluation of teaching in American colleges and universities.... A judicious evaluation of a college professor as teacher should include: (1) an accurate factual description of what an individual does as teacher (AAUP Committee C , 1975).[11]

How SEF establishes incompetence, or neglect of duty is problematic, having largely to do with issues of validity. To further complicate matters the concept of academic freedom, like most abstract terms is logically fuzzy around its edges. Moreover, unlike the legal categories of academic freedom and tenure, there is no equivalent legal category of SEF. Consequently, published legal rulings on this issue are scarce.

While academic freedom has not been recognized universally by the courts as equivalent to a constitutional right, it has nevertheless been viewed as a right which the courts have deemed must not be violated in the performance evaluation process. In addition, academic freedom has been associated with the First Amendment right of free speech. Some courts have considered it to be a First Amendment-right in and of itself. Academic freedom is a "special concern of the First Amendment, which does not tolerate laws that cast a pall of orthodoxy over the classroom." While the two rights are not necessarily the same, they frequently and sufficiently overlap to trigger judicial scrutiny when faculty performance evaluation process threatens to impinge on the First Amendment (Copeland and Murry, 1996).

Faculty Assessment of How SEF Infringe Upon Academic Freedom

Mention of this issue within the text of a number of articles throughout the literature provides a kind of uncontrolled data base indicating that an increasing number of faculty consider SEF an impingement on academic freedom. There are, however, a few formal surveys. While the judgement of a profession's membership on a subject is not necessarily a valid indicator, especially when that the judgement renders them the primary beneficiaries, in the absence of significant evidence to the contrary it would seem reasonable that a profession's judgement should be accorded considerable weight in any analysis or process leading to policy decisions.

Before reviewing what the literature seems to show is the predominant faculty view, a statement of at least one faculty who does not think SEF impinges on academic freedom will illustrate the opposite perspective. This faculty was quoted as saying,

> Do I sense more than a little self-disdain in the comments of my colleagues who portray faculty members as cowed by the prospect of a less than enthusiastic course evaluation by their students? That is not my impression of my colleagues, and I can assure you it is not an accurate representation of me. It is my impression that faculty grade students with quality of performance as the primary, if not sole, criterion (Kresl, 1995).

At least some faculty, then, believe that SEF does not influence their behavior and is therefore not an abridgement of academic freedom.

While formal surveys of faculty views on this issue are relatively rare, one study found that at least one third of faculty respondents reported lowering their grading standards and course level in response to their student evaluations (Ryan, Anderson, and Birchler, 1980). Another study found that 64.8% of respondents believed that "student evaluation forms are responsible for lenient grading." According to another study, 39% of accounting administrator respondents admitted being aware of faculty who altered their instructional behavior in order to improve evaluation scores (Crumbley and Fliedner, 1995). Faculty were also in nearly universal agreement that SEF is important in promotion (86.6%) and tenure (88.2%) reviews (Kolevzon, 1981, see also Avi-Itzhak and Lya, 1986).[12] At the very least, such reviews are strong prima facie evidence of SEF as an infringement on academic freedom.

In two nationwide studies of accounting departments it was shown that 37% of faculty respondents were dissatisfied with their present evaluation system. The response of 561 accounting faculty to the statement that the present system of SEF is well-designed and properly implemented, reported that 15.3% indicated "strong disagreement," 31.7% "disagree," and 25.5% indicated a "neutral" response (Yunker and Sterner, 1988; Bures, DeRidder and Tong, 1990).[13]

Unlike the paucity of formal surveys, there are numerous statements by faculty in the research literature clearly arguing that SEF is an infringement on academic freedom. These statements by faculty contend that SEF (1) is prime facie evidence of administrative intrusion into the classroom, (2) are often used as an instrument of intimidation forcing conformity to politically correct standards (Young, 1993), (3) create pressure for a self-policed lowered teaching standard (Bonetti, 1994), (4) are responsible for a considerable amount of grade inflation (Greenwald, 1996, Greenwald and Gillmore, 1966),[14] (5) function as prescriptions for classroom demeanor (Damron, 1996), (6) when used for promotions, salary raises or continued employment, SEF becomes a potent means of manipulating the behavior of faculty (Stone, 1995), (7) when salary and promotion are possible consequences of SEF there is pressure for faculty to teach in a manner that results in higher student evaluation (Damron, 1996), (8) contrary to their original intent of improving instruction, do not eliminate poor or below-average teachers but instead increases poor teaching practices (Carey, 1993), (9) illustrate a mercantile philosophy of "consumerism" (Benson, and Lewis, 1994), which erodes academic standards (Goldman, 1993; Renner, 1981), (10) have thus lowered the quality of U.S. education (Carey, 1993; Crumbley, and Fliedner, 1995; Young, 1993), (11) lead to the inappropriate dismissal of faculty (Parini, 1995), and (12) constitute a threat to academic freedom (Dershowitz, 1994; Stone, 1995). Finally, it would seem that SEF creates an educational conflict of interest between faculty and student impacting on the quality of instruction.[15]

At least one Faculty Senate tenure and promotion committee at a large state university along with the ACLU (Heller, 1986), found that given the crude state of SEF and its subsequent use, such evaluations are a"de facto violation of academic freedom." (p.14) Despite these concerns, however, there are some faculty who, given the alternative of peer evaluation, consider SEF as less harmful (Carey, 1993).

Academic freedom involves more than just external political control of course content (Schrecker, 1986), and a Scopes Trial type gag order; it entails pressure to censor unpopular subjects within the popular cultural belief systems of students. In this regard, Alan Dershowitz (1994), the well known Harvard Law School professor, and appellate attorney believes that the administrative use of SEF is a direct threat to academic freedom and to quality education. After teaching on a controversial subject, and receiving negative feedback from students, he says he "realized how dangerous it would be for an untenured professor" to teach about such subjects. He goes on to point out, however, that "Most of the students appreciated the diversity of viewpoints," and evaluated him fairly, noting that he was "very good at presenting alternative views," and "helped me get a less dogmatic view of the law," that he was "open to criticism," that the class was "the most engaging class on campus," that he was "the most intellectually honest professor I've had," and was "fair in presenting sides that usually aren't raised."

There was, however, a sizable group of students who, says Dershowitz, "used the power of their evaluations in an attempt to exact their political revenge for my politically incorrect teaching." He was lecturing on the (sensitive) legal analysis of rape.[16] He notes, that

> One student said that I do "not deserve to teach at Harvard" because of my "convoluted rape examples." Another argued that women be allowed an "option" not to take my class because I "spent two days talking about false reports of rape," another demanded that my "teaching privileges" be suspended. One woman purported to speak for others: "Every woman I know in the class including myself found his treatment of rape offensive and disturbing." Another woman felt "oppressed throughout the course." Although I always try to learn from my evaluations, I will not be bullied into abandoning a teaching style that I believe is best designed to stimulate thinking.(italics added, p.118-119)

Dershowitz says that it takes no courage for him to exercise his academic freedom, since he holds the rank of full professor and has tenure, but he seriously wonders if he were an untenured assistant professor, if he would I have the same courage. He concludes:

> Are other less established teachers being coerced into changing their teaching by the fear of negative evaluations, which can be fatal to tenure? You bet they are, and it poses a real danger to academic freedom and good education. (italics added, p.118-119)

Dershowitz perhaps overstates his case on the security provided by tenure for faculty on many campuses. Being at Harvard, he is not the typical faculty member. Unlike at the Harvards, the Seven Sisters, and the Ancient Ones, in the trenches of most campuses SEF is accorded inordinate weight (at least by administrations) over research. He is also "politically" protected by being so well known, not to mention that his famous cases and books have probably made him financially independent.

Classroom and Curricula Infringements Based on SEF

The following three recent examples will illustrate the extreme to which SEF can lead in terms of impinging on academic freedom. Each exemplifies a different aspect of academic freedom being infringed upon by SEF as used at most institutions. There are likely a number of cases similar to a recent one reported in the Chronicle of Higher Education of faculty being dismissed because of negative student evaluations (Magner, 1995).

With four weeks remaining in the semester, a non-tenured faculty was removed as the instructor of an introductory chemistry course at the University of Montana.[17] His dismissal was prompted by a student petition signed by 100 of the 200 students in one of his classes. These students maintained that he demonstrated an "inability to teach new information" and a "willful academic demoralization of students." Three different views of this situation have emerged. Some contend that he was a tough teacher who became the victim of "whiny" students looking for an easy instructor to spoon-feed them the course material. Supporters maintained that many students in the course planning professional or graduate programs, feared that a low grade in a difficult subject would diminish their chances of admission. Others, maintained he ignored student questions, and made tests more difficult than they needed to be.

The faculty member maintained he received good student evaluations in the past with no problems in his other courses. The faculty believe the reasons for the complaint is that he is a demanding teacher who requires students to go beyond simple memorization and regurgitation of facts on his exams. He says he has "never knowingly belittled a student." He sees an anti-intellectualism at work, maintaining that "There is an attitude on...campus that any display of learning or erudition is arrogance." He also maintains that "it's not the students' prerogative to decide who teaches a course, and by extension, how that course should be taught." Administration is quoted as saying, "We're an open-admission university. A very large fraction of the class was completely unable to compete." As it turns out, the previous year the administration warded off a similar student petition against a tenured full professor of chemistry. In that case, the administration says they "clearly felt we were dealing with some people [students] who wished to get better grades for less work." Granted, all the facts are not in on this case. The faculty was replaced with a professor who is popular, with students. Nevertheless, it is clear that student evaluations were the primary mechanism leading to a faculty's dismissal when, whatever the facts are, they are at least open to serious question.

As well as academic freedom in the classroom, other cases affect tenure, promotion, and more importantly, curricula. A faculty (Goldman, 1993) at Wichita State University notes that in a thirty-faculty department which is responsible for certification of teachers, in the past 25 years, six faculty have been hired as assistant professors to teach Foundations of Education. All apparently well qualified, receiving their doctorates from excellent universities, only one of these faculty was awarded tenure; none was promoted. According to Goldman (1993), the reason for this is student evaluations.

In general, as the data show, required courses hold less interest and receive lower evaluations than elective courses; in general, students who are drawn to become teachers are concrete-sequential, and are less interested in the abstract and theoretical content of the foundations of education course. This leads the faculty who teach the foundational course to receive lower student evaluations than other education faculty. Because student evaluations are often the major, if not the only, gauge of teaching quality, and since teaching quality usually out ranks research and scholarly productivity on most campuses, when tenure, promotion, and salary increases are awarded, these rewards will not be evenly distributed to the foundational, educational psychology faculty. Rewards will accrue to the concrete-oriented methods faculty whose courses will further intensify the concrete orientation of teacher preparation. And the downward spiral will continue.[18]

The third example of SEF as infringement on academic freedom reportedly occurred in a business department (in Crumbley, 1995). A rigorous instructor teaching a basic business class gave D's and F's and received SEF scores in the one range (on a 5-point scale). As a consequence, she was removed from teaching that class and assigned to a non required graduate course where she proceeded to give a 50-50 split of As and B's each semester. She received SEF scores of up to 4.9 after making this adjustment. She informs the students at the start of the semester that only As and B's will be given. This administrative strategy of assigning "tough" grading faculty to non required courses allows students to force easy grading by self-selecting away from the more stringent graders, and thus "censors" certain instructors courses and by implication the content of what they teach.

These examples suggest how SEF can be used to shape faculty behavior, curricular content, and the kind of faculty that are retained in programs. They also suggest how SEF is selectively used as a primary mechanism of dismissal of faculty who do not conform to student and administrative demands. Though a faculty may (a) receive good evaluation in other courses, and (b) receive good evaluations from the majority of students, a few negative evaluations may be determinative of disciplinary action toward faculty. The examples also raise the issue of who and what determines academic standards. On campuses with "open admissions," meaning a lower level of student in courses than has historically been the case, that an otherwise competent instructor is required to adapt his or her course to whatever level of student is enrolled, or be subject to the above consequences based on student dissatisfaction. Regardless of the specifics, the case illustrates the direct aspects of student evaluations affecting educational standards and raises the question of academic freedom of faculty in the classroom.

Legal Considerations of SEF

There are multiple latent legal issues engendered by SEF. Unlike academic freedom and tenure, as a legal search category SEF does not exist.[19] In a subsequent paper I will coalesce and deal more fully with legal rulings involving SEF in the denial of reappointment and tenure. While the following ruling is Canadian and has no standing in the U.S., it is nevertheless initially relevant, especially given that Canadian protection of free speech is more narrow than in the U.S.

Under Canadian law, a situation recently developed at the University of Regina where a faculty was denied tenure partly on the basis his student evaluation scores (Education Employment Law News, 1994).[20] A board of arbitration ruled that

> tenure decisions could not be based solely on assessments which were completed by students who had never been made aware of the ramifications of their statements....To

> base serious career decisions narrowly on student evaluations is not to be encouraged..... This is particularly so when...the students are not advised of the potential use of the evaluation tool... [I]f evaluations are to be used for serious career development purposes those completing them should be aware of the potential consequences of their participation.

At least one similar ruling exists under U.S. law.

At the University of Guam, a ruling to remove anonymous student evaluations from professors' tenure files was handed down by an arbitration board as the result of a rare challenge to the use of such evaluations in tenure and promotion decisions (Blum, 1990). The action was in response to a grievance filed by the university's faculty union, the Guam Federation of Teachers, which is affiliated with the American Federation of Teachers (AFL-CIO). The union said the use of SEF violated the union's contract with the university, which provides that anonymous documents or those "based on hearsay" should not be included in a faculty member's file. The union also argued that the university improperly interpreted the data from the evaluations.

Some of the issues here are (1) students not being made aware of the purpose and ramifications of their evaluations, (2) the anonymous nature of student evaluations, (3) the invalid analysis of SEF, and therefore, (4) SEF in effect being anecdotal and hearsay data. Since most SEF results are prepared anonymously, an instructor has no recourse to confront his/her evaluators. As will be addressed below, the anonymous nature of SEF is beginning to also be questioned by arbitration boards.

After reviewing the research, a Director of the Office of Educational Assessment at one large state university (Gillmore, 1984) concluded that, "If student rating are to qualify as evidence in support of faculty employment decisions, questions concerning their reliability and validity must be addressed." (p.561) Perhaps the most comprehensive review and statistical analysis of SEF validity is that of Greenwald (1996).

The philosopher of science, Michael Scriven (1995, 1993, 1991), who has conducted extensive analyses on faculty evaluation methodology suggests of faculty evaluation in general that,

> All are face-invalid and certainly provide a worse basis for adverse personnel action than the polygraph in criminal cases. Based on examination of some hundreds of forms that are or have been used for personnel decisions (as well as professional development), the previous considerations entail that not more than one or two could stand up in a serious hearing.[21]

The cost of lawsuits notwithstanding, given the above, it is surprising that there have not been more lawsuits by faculty.[22] It appears, however, that in regard to faculty evaluation in general the courts have not been concerned with validity issues, though this may be changing as well (Rebell, 1990).

Releasing SEF to Students and the Public

In exploring possible legal implications of SEF, it should be made clear that I am not an attorney and approach this section on the basis of the "reasonable man" legal standard. To begin, some faculty believe that due process and defamation issues are involved in SEF (Crumbley, 1996). It has been suggested that faculty are entitled to at least the same rights as students. The Fourteenth Amendment requires, for example, due process before a public institution may deprive one of life, liberty, or property. Given the problematic nature of SEF, due process is in question. In a university, a faculty's reputation is considered a liberty right, and for tenured faculty the courts have pronounced the possession of tenure a property right. Presumably, any inappropriate action depriving faculty of these rights would be open to legal action.

Though it is illegal to post a student's grades using a social security number or date of birth,[23] on the majority of campuses scientifically questionable SEF and other anecdotal student remarks about faculty teaching are not only used in determining faculty salary increases, promotion and tenure decisions, they are openly published on some university campuses and sanctioned by some administrators and state government officials. In what many faculty see as an outrageous attempt to control the academic classroom, some state governments have sanctioned the release of SEF to the campus community and in some cases to the general public by publishing faculty student evaluations on the university's world wide web pages, thus making them not only available on campus but globally.

Case in point: At the University of Wisconsin, the Chancellor refused to release the SEF, citing a statute allowing personnel evaluations to be withheld from public view. The students took the chancellor to court. However, after being advised to do so by the state's Attorney General, citing Wisconsin's open-records law, the University of Wisconsin's campus will open students' evaluations of professors for public view. To the credit of the student and faculty senates, they passed resolutions in support of the Chancellor's refusal, and the university's lawyer concurred. Despite these resolutions, the Attorney General disagreed, writing that "the requested records are public records and the university's stated reasons for withholding access do not outweigh the public interest in the records" (*Chronicle of Higher Education*, 1994a, 1994b).

Other schools also published SEF. One recent survey of accounting departments found that 11.4% of the respondents indicated that SEF scores are made available to students (Crumbley and Fliedner, 1995). Indeed, a search using "faculty evaluation" on the world wide web will return numerous examples of published SEF. All this while faculty are restricted from divulging information on students (see Pennsylvania State University, 1996). Articles are, however, beginning to appear that question the legality of publically releasing SEF (Robinson and Fink, 1996).

It has been suggested that if a university damages a faculty's reputation by publishing false and anecdotal data from SEF, faculty should able to sue for libel or defamation. The concept of defamation typically refers to communication that causes a person to be shamed, ridiculed, held in contempt by others, or their status lowered in the eyes of the community, or to lose employment status or earnings or otherwise suffer a damaged reputation. Legally, while defamation is governed by state law, it is limited by the first amendment (Black, 1990).[24] According to one source, however, the courts have generally protected administrators from defamation charges resulting from performance evaluations (Zirkel, 1996). It would seem, however, that these older precedents applied when administrative evaluations were conducted in private and not publically distributed.

University administrators are often allowed to release SEF to students when the release of personnel information is apparently allowed in no other phase of personnel or other key management functions. An Idaho ruling upheld the release of SEF to students by reasoning that students were not the general public and therefore faculty evaluations were not protected under the privacy rights of the Idaho Code (Evaluating Teacher Evaluations, 1996). Given such apparent breaches of confidentiality and privacy, it will be instructive to see how the courts will continue to rule. It would seem that a university should be held responsible for insuring that data made public are valid.

Finally, in typical personnel evaluations, professional validation studies are not permissible unless shown by professionally acceptable methods to be "predictive of or significantly correlated with important elements of work behavior which comprise or are relevant to the job or jobs for which candidates are being evaluated." In Title VII of the Civil Rights Act of 1964 the employer must meet "the burden of showing that any given requirement (or test) has a manifest relationship to the employment in question" (in Griggs v. Duke Power Co., 401 U.S. 424 (1971). "In view of the possibility inherent in subjective evaluations, supervisory rating techniques should be carefully developed, and the ratings should be closely examined for evidence of bias" (EEOC Guidelines, 99 CFR 1607.5 (b) (4). (in Crumbley, 1996).

SEF As Social Judgement and Diagnosis

Given courts uncritical assumptions regarding the validity of faculty evaluations in general, and the untrained judgement of those making decisions, it would seem appropriate to address the issue of judgement and decision making in SEF. Frequently, in interpreting SEF the method is simply "eyeballing" them. The "actual" consistency of findings of SEF validity at this point in time, notwithstanding, one judge (Rebell, 1990) clearly points out, "because the state of the art concerning teacher-evaluation practices is at a sensitive developmental stage, extensive court intervention at this point can substantially influence ---for better or worse---the future direction of basic practice in the field." (p.344) He goes on to say that seen in this light, it is clear that whether increased judicial intervention will have a positive or a negative impact on professional evaluation practice lies more with educators and psychometricians than with the judges, meaning that faculty control over future rulings depends upon providing clear evidence to the courts.

Thus, part of influencing future court rulings is demonstrating relevant research. With the typically unsystematic method by which SEF are interpreted, and the considerable weight accorded selective negative comments by a few students in making tenure and promotion decisions, the meaning of SEF (i.e., diagnosis) as they are applied to the analysis of teaching effectiveness is, in fact, all too frequently an intuitive one.

As to the issue of the validity of findings in SEF, as indicated at the opening of this paper, more recent and rigorously analyzed statistical findings cast serious doubt on their validity. The research literature on judgement and decision making would seem applicable to both students making such judgements and to those interpreting (i.e., diagnosing) the results, especially given the intuitive and non-data based foundations of such judgements and decisions. In the course of reviewing the research over the years on social judgement and clinical diagnosis, it seems clear that the manner in which nearly all SEF data are analyzed is a subset of the research on social judgement and clinical diagnosis, with both student and interpreter being involved in the same logical and cognitive biases and distortions that result in the pervasive low accuracy level of social judgement in general and clinical diagnosis in specific.

Psychological research has recognized the severe cognitive problems and limitations of "intuitive," and "experience-informed" everyday judgements for over thirty years, (Dawes, Faust, and Meehl, 1989; Faust, Guilmette, Hart, Arkes, Fishburne and Davey, 1988; Garb, H. N. 1989; Hayes, 1991; Larkin, McDermott, Simon, and Simon, 1980; Rabinowitz, 1993) yet the mistakes continue in everyday practice situations. Interpretation of SEF are no different. As Franklin & Theall (1990) independently observe relative to SEF

> Even given the inherently less than perfect nature of ratings data and the analytical inclinations of academics, the problem of unskilled users, making decisions based on invalid interpretations of ambiguous or frankly bad data, deserves attention. According to Thompson (1988, p. 217) 'Bayes Theorem shows that anything close to an accurate interpretation of the results of imperfect predictors is very elusive at the intuitive level. Indeed, empirical studies have shown that persons unfamiliar with conditional probability are quite poor at doing so (that is, interpreting ratings results) unless the situation is quite simple." It seems likely that the combination of less than perfect data with less than perfect users could quickly yield completely unacceptable practices, unless safeguards were in place to insure that users knew how to recognize problems of

> validity and reliability, understood the inherent limitations of rating data and knew valid procedures for using ratings data in the contexts of summative and formative evaluation. (79-80).

The authors conclude by noting "It is hard to ignore the mounting anecdotal evidence of abuse. Our findings, and the evidence that ratings use is on the increase, taken together, suggest that ratings malpractice, causing harm to individual careers and undermining institutional goals, deserves our attention." (pp. 79-80). Recognizing such problems is not methodological nit-picking; they are pragmatic and fundamental.

It would therefore be a reasonable recommendation to suggest those researching the validity and especially interpreting SEF familiarize themselves with the psychological social judgement and clinical diagnosis literature. Few states or courts require evaluators to be trained in evaluation methods (Rebell, 1990).[25] It is these cognitive problems that need to be addressed in the SEF literature prior to otherwise statistically sophisticated of the data being performed analyses---when rigorous methods are used.

SEF and Administrative Control of Academic Freedom

In addition to legal aspects, there are the pragmatics of the SEF which maintain its use. The literature clearly suggests that administrators tend to strongly oppose the elimination of SEF being used for faculty salary, promotion and tenure decisions. There are three reasons for opposing the elimination of SEF. The first seems to be a lack of practical alternatives to SEF (Greenwald, 1996),[26] the second is administrative control, and the third is that student input facilitates student retention in numerous ways.

SEF provides a mechanism of control in a system otherwise lacking direct control over faculty, and are a powerful tool in assuring classroom changes that lead to the retention of student tuition dollars by assenting to student consumer demands and of parents who foot the tuition bill. Academic issues such as teaching, grading, curricular requirements, and other academic standards have by tradition and expertise been the exclusive province of faculty. Although academic freedom and the protection of tenure would appear to insure faculty classroom independence, the extent to which faculty thought and behavior are administratively shaped is the extent to which both are infringed upon (see below).[27] Such administrative control mechanisms, of which SEF is one, are therefore seen as infringements on academic freedom.

Control mechanisms are more widespread and intricately embedded in the everyday operations of the university than is generally acknowledged. After a review of the research, Stone (1995), observes that SEF opens the "door to the direct application of bureaucratic control to academic decisions. It is the very kind of policy that, for example, has enabled educational administrators to mandate the "politically correct" at the expense of the "academically credible." He further observes,

> In great part, the reason for this outcome is that the faculty comprising colleges, departments, and programs are subjected to the same incentives that impact the institution as a whole. Chief among these incentives is internal flow of institutional resources. The allocation of institutional resources is administratively controlled in such a way that the academic units with the greatest enrollment growth are generally afforded the greatest resources. The incentives are such that no individual faculty member can comfortably question activities that build his or her department's or college's budget. After all, the job they save may be their own. The same can be said for peer evaluations of faculty for promotion, tenure, and merit pay. Faculty typically find it difficult to negatively evaluate colleagues who are high credit hour producers

> and who have high student ratings even if their academic standards are suspect. Credit hours and student ratings are routinely and carefully monitored by the administrative bureaucracy. Learning outcomes are given much less attention. Units that fail to grow not only remain under supported, they are sometimes cannibalized, and their share of the budget is channeled to the areas demonstrating greater potential for growth....The internal competition for resources shapes individual faculty behavior as well.... They are, therefore, in the interest of self-preservation, especially inclined to accommodate academic standards to the student market.

Stone concludes that administrative influence, including the use of SEF, has become so effective in shaping faculty that their collective voice has been seriously compromised.[28] It should be noted that some administrators, and university presidents do recognize the importance of tenure and support its continuance. (e.g., Cotter, 1996).

Contrary to some published reports, conditions such as weakened standards, fragmented curriculum, and inflated grades do not simply arise from a spontaneous deterioration of faculty into so-called "deadwood." Neither are most faculty incompetent at teaching. As Stone (1995) points out, "Rather, these problems seem likely to have developed as a result of the continuing insidious pressure placed on teaching and grading practices by the imperative to keep students happy and enrollments up." While most students are positively disposed to SEF, not all students are so oriented.[29]

Many students understand the above described ensuing consequences. A glance at articles from online student newspapers reveals strong sentiments against what many students consider the erosion of standards created by SEF. One student writer went so far as to say "We therefore suggest a boycott of the 1995 student/teacher evaluations. This boycott will provide a more effective means of communication than anything written on the evaluation itself. Something must be done about the trend of grade inflation. We as students refuse to contribute to the downfall of academia " (Stern and Flynn, 1995). Some students are thus quite aware of the effects of SEF on their education. Thus the foregoing are something to seriously consider in the light of the principle of academic freedom.

Faculty Complicity and Adaptation to SEF

There is the argument that academic freedom is not abridged if faculty voluntarily participate in the SEF process. This argument is problematic for three reasons. First on most campuses SEF is required by administrative policy, and is therefore neither individually nor collectively voluntary.[30] Second, as noted by Stone (1995) above, given their mandated use in salary, promotion and tenure decisions, SEF creates significant pressure to conform to administrative and student demands. The third is that the second condition does not seem to meet the requirements of what is legally considered "informed consent" in a contractual agreement.

Like most people, given that faculty behaviors are subject to reward structures, is it is not surprising that faculty would tend to teach to the evaluations and self-limit their academic freedom prerogative. From a purely psychological/behavioral learning theory perspective, when a faculty's livelihood is at stake, it is understandable why many will adjust their grading and course content level in order to receive a good evaluation. It is also predictable: behaviors that are rewarded tend to increase the probability of their reoccurrence.[31]

Similarly, as two research studies note (Crumbley, 1995; Nelson and Lynch, 1984), from an economic business accounting point of view, If a faculty can choose teaching styles, grading levels, and course content, s/he will naturally prefer choices that are expected to result in higher SEF scores; if faculty know the variables affecting their careers, they will meet these criteria. There are laws that regulate ones financial statements so as to reduce the manipulation and opportunistic behavior regarding income, yet there is no regulation of SEF. Faculty "have a high incentive to manage SEF, even more so than managers have the incentive to enhance earnings"(Crumbley, 1995). Unfortunately student ratings tend to discourage instructors from espousing views that might offend popular student prejudices. "Expecting that even highly professional individuals will disinterestedly adhere to academic and intellectual ideals in the face of pervasive incentives to do otherwise is not realistic." (Stone, 1995)

Even one of the strongest advocates of the validity of SEF (Seldin, 1984) warns: "The confidentiality of the data must at all cost remain inviolate. If data are shared, it must be with the consent and at the discretion of the appraised professor. For data surreptitiously to be used for personnel decisions, it will have an immediate chilling--even fatal--effect on the credibility of the entire evaluation program." (p.129) Certainly when SEF are published not only confidentiality has been violated but so is trust. Again, given the above pressures, it would seem that "voluntary" faculty participation in the SEF process does not meet the requirements of what is legally defined as informed consent. As two legal scholars (Copeland and Murry, 1996) noted in a different context is directly applicable here. They note, "Of course the right to academic freedom means very little if by exercising it the educator suffers, or runs the risk of suffering, financial consequences." (p.249)

SEF As Trojan Horse In The Movement to Eliminate Tenure

There is a final and even more serious threat that SEF can be said to pose to academic freedom than the simple act of student pressure and administrative control of rewards based on them. This last threat is not generally recognized. This threat is the use of SEF in a continuing national movement against tenure and to reduce costs. For about the last twenty years there has been considerable criticism about the concept of tenure. The public has always viewed the purpose of academic tenure as job security, just like seniority in a trade union. And as with unions, the perception is that job security leads to lowered productivity. Accordingly, there have been continuing periodic attempts to eliminate tenure. More recently, however, these attacks have increased. The attacks come from all levels of society and from inside academia itself, including boards of trustee and State Legislatures attempting to either eliminate or revamp academic tenure, the latest being the widely publicized bid by the University of Minnesota Board of Regents (Guernsey, 1996; Healy, 1996; Magner, 1996; *The Chronicle of Higher Education*, 1995).

Typically, the arguments against tenure have been economic. With shrinking budgets, these old economic arguments have taken on an even more bottom line approach to education, arguing, at the least, for the revamping of tenure for the purpose of enabling schools to become programmatically and fiscally more flexible in a changing world. While views of tenure are changing, including those of the AAUP, traditionally AAUP has primarily held that "Institutions of higher education are conducted for the common good and not to further the interest of either the individual teacher....[and]..The common good depends upon the free search for truth and its free exposition. Academic freedom in its teaching aspect is fundamental for the protection of the rights of the teacher in teaching and of the student to freedom in learning" (AAUP, 1970).

At least one court has ruled, "The purpose of tenure is to protect academic freedom---the freedom to teach and write without fear of retribution for expressing heterodox ideas"(in Copeland and Murry, 1996, p.250).[32] Herein lies the point: academic freedom can only be guarded within a system of tenure.[33] Certainly tenure does not guarantee academic freedom. Nevertheless, though tenure is not a sufficient condition, it is a necessary condition for academic freedom.

Some tenure system makes it procedurally difficult to fire faculty who may disagree with popular beliefs. There is a long history, continuing to the present, of professors being fired because of their unpopular critiques. The classic examples, of course, include the 1925 trial of John T. Scopes, the Tennessee high school teacher who was fired for teaching the Darwinian theory of evolution. In the 1950s, we saw Senator Joseph McCarthy's communist witch hunts that resulted in professors being fired and/or blacklisted. During the 60's and 70's entire departments were wiped out or put in special trusteeship because of faculty critiques of the Viet Nam war. More recently, "political correctness" about gender, race, and other issues, both on the left and the right, has led to the firing of professors, both tenured and non tenured. So even tenure is not an absolute guarantee for academic freedom.

Part of maintaining safeguards to academic freedom is that the university be insulated (not isolated) from the popular culture. To understand tenure, then, it is necessary to understand that the university is a unique organization and in its most important aspects can not be compared or likened to other organizations in society. First, unlike the business corporation, the university should not be consumer oriented (though there are, of course, certain economic realities). The role of the university is leadership, not as a servant of consumer (i.e., student) demands.

The university tenure function insures one of the only places in society where open dialogue on any issue no matter how unpopular or unorthodox can be critically examined without consideration of the political cost, without fear of reprisal, without the pressures of social taboos, social norms, faddish movements, personal notions of etiquette, and other immediate pragmatic pressures that exist in the culture in which it is embedded. Above all other roles, this is the defining feature of the university in a democratic society. Without this role, there is no place where truth--- with either a capital "T" or lowercase "t"---can be sought. Once the university becomes indistinguishable from the popular culture, or becomes political, it ceases to be a credible place for the creation and dissemination of knowledge.[34]

If tenure is compromised, we will no longer have a university as we know it. It may look like a university, but it will not in fact be one.[35] It will be something else. If we want to play chess without using the King, we can certainly elect to do so, but whatever game we are then playing, let us be clear that it is not chess.[36]

Undergirding this push to eliminate or revamp tenure lies a metaphor that has been transformed into a literal concept: The university as a business entity. The university considered as a business carries with it the attendant and associated ideas of students as consumers in an educational market place, fiscal efficiency, and being a servant of the consumer community in which it is physically embedded. (See Sommer, 1995). The fact is, a university is not like a business. Some courts have recognized inappropriateness of this metaphor. One has noted (EEOC V. Franklin & Marshall College, 1985)

> I do not agree with the majority's assumption that academic institutions are the same as any other employer. At least insofar as their administrative and governance structures are concerned, colleges and universities differ significantly from garden variety private employers. In the context of application of the provisions of the National Labor Relations Act the Supreme Court has counseled that "principles developed for use in the industrial setting cannot be imposed blindly on the academic world.'" NLRB v. Yeshiva University, 444 U.S. 672, 681, 100 S.Ct. 856, 861, 63 L.Ed.2d 115 (1980) p.120

Academic tenure, then, is not like seniority in business, civil service, or a union where the purpose of "tenure" is for the protection of the worker. Unlike these organizations, academic tenure is for the protection of the education function, not individual faculty jobs; it is for the protection of the role of the university in a democratic society. In fact, the metaphor of student as consumer is more appropriately replaced by the metaphor of student as worker or apprentice. Enter SEF.

SEF has become the newest weapon in the attack on tenure. But the assault is not a frontal one. SEF is increasingly used in evaluation of faculty for tenure and low evaluation is taken as prima facie evidence of teaching ineffectiveness, thus justifying not awarding tenure to a faculty. A recent legal handbook for college administrators (Weeks, 1996)says, "Student ratings are an effective tool to respond to litigation...[and]...They can become important evidence if a college should decide to terminate a tenured professor for reasons of competence." Of course findings from SEF may indeed validly describe incompetence, but given (a) the conflicting data on their validity, (b) the way many institutions have constructed SEF instruments, (c) the often unsystematic statistical method by which SEF are interpreted, and especially (d) given the considerable weight accorded negative comments by only a few students in making tenure and promotion decisions, it would seem SEF can all to easily be used as a covert instrument for the elimination of tenure candidates and other faculty who may threaten student tuition dollars and perhaps ideological and popular culture agendas.

As noted above SEF can be a powerful administrative control mechanism on granting tenure. Abolishing tenure also lowers operating costs. Older faculty who may be unpopular, and who lost more, can be fired and younger, cheaper, faculty can be hired. It is often older faculty who consistently receive lower student evaluations than younger faculty (Feldman, 1983). McMurtry (1991) has noted that, education has always been subject to external pressures whose purpose is to subordinate it to vested interests of various kinds, whether it is slave-holding oligarchies, theocratic states, political parties or the prevailing dogmas of collective beliefs. The difference today, however, is that threats to academic freedom come from within: the consumer student.[37] As the well known scholar David Reisman (1981) noted years ago in this regard "This shift from academic merit to student consumerism is one of the two greatest reversals of direction in all the history of American higher education; the other being the replacement of the classical college by the modern university a century ago." (p.xi) Within a consumer model of education, to deny consumer demands is difficult. While it is difficult enough to deal with political, ideological and economic pressures, dealing with consumer pressures has become nearly impossible. If denying fiscal efficiency is viewed as unreasonable, irresponsible, and even irrational, to deny "consumer's" their demands is viewed as undemocratic and downright mean spirited.

It might be asked that if tenure is not joined at the hip with academic freedom, why are Supreme Court Justices granted lifetime tenure? It might be asked, too, why it is that virtually no reasonable person suggests revamping academic freedom, yet many have little qualm--if any---about eliminating or at least about revamping tenure? The short answer is that for many tenure has come not to be seen as crucial to academic freedom. The longer answer is that at least in principle, revamping academic freedom would not be acceptable in a democracy. At worse, it would be viewed as the equivalent of Gestapo busting in the front door of the university (or as a kind of Tiananmen square of higher education). What is not widely understood is that SEF is often a kind of Trojan Horse in the battle against tenure and academic freedom.[38] It often becomes a stealth mechanism by which to covertly abrogate both tenure and academic freedom.

SEF and Academic Freedom in the 21st Century

As noted above, arguments against tenure have typically been economic ones. As higher education enters the 21st century, and its associated demographic changes, however, arguments against tenure are changing. A paradigm shift is taking place in arguments against academic freedom and tenure, a paradigm that is based in the changing demographics of the student population. It is said that academic freedom tends to be viewed from the perspective of a bygone era when the university faculty and student population were relatively homogeneous. Accordingly, eliminating tenure or at least radically revamping it is increasingly being justified not on matters of principle but by political and other expedient considerations.

It is currently suggested that eliminating or revamping tenure would be fairer to minorities, to unemployed, and to part-time faculty, (Murray, 1996; Parini, 1995; Wilson, 1996) presumably by opening up faculty positions for young professors and minorities. Thus critics of tenure maintain that a university going into the 21st century can no longer afford the luxury of the traditional view of academic tenure. While these are valid problems changing the parameters of academic freedom and tenure is a dangerous way to solve them. To the contrary, in a 21st century global society, academic freedom and tenure become more important than it has been in the past, not less. This is why:

With increasing diversity in both the general culture and the university campus, including not only diversity among the student body, but diversity among the faculty, comes increasing conflict of ideas, values, and perspectives. To manage these changes, increasing attempts to reduce the conflict by limiting freedom of speech and action in the name of protecting minorities are being implemented. For example, it has been charged (Schrank, 1993) that,

> As now defined, academic freedom...ignores the intersubjectivity of all persons in the setting... such an approach conceals the vulnerability of women and other historically excluded groups who are still marginal in the academy, and does not take account of the historic advantages enjoyed by white, heterosexual, able-bodied males...it does not acknowledge power imbalances in relations based on gender, race, sexuality, class, and other dimensions of difference....An exclusive focus on words and ideas can obscure the fact that we are talking about whole people. For many people, ideas and words may well be referents for highly significant experiences which have powerful meanings, evoke strong emotions, and are not simply ideas and words. The meanings and emotions attached to ideas and words render people extremely vulnerable in what, for others, might be a benign intellectual exercise....The argument for including civility recognizes that ideas are not separate, public, and objective but real, emotional, and personal.

Certainly, insensitivities in the classroom are to be discouraged. What this quote reflects is a fundamental paradigm shift in the parameters of academic freedom as historically conceptualized. SEF can and do reflect these and other political and cultural conflicts, creating what the courts in other contexts have called a "chilling" effect on academic freedom (Recall the above emotional evaluation by students described by Allan Dershowitz).

This paradigm shift in the parameters of academic freedom, however, is but a subset of a more overarching social shift in the first amendment right to free speech outside of academia. In an extensive article in a noted law review journal (Matsuda, 1989), the author opens with the statement, "This article rejects an absolutist first amendment position." (p.2321) Such views argue for restricting academic freedom on the basis of the feelings of various and changing constituencies. The problem, as Dershowitz (1994) points out, "is that everyone has a different...exception, and in a nation of equal protection, it is difficult to pick and choose among the proffered exceptions. If we were to accept them all, there would be little left of the First Amendment." (p.41) In academia, this means there would be little left of academic freedom and quality education.[39]

Contrary to popular views, tenure and the traditional concept of academic freedom has not simply protected the majority, it has and will protect the newly entering minorities. Indeed, women and other minority studies programs in the university largely owe their very existence to the traditional principle of academic freedom. As Nadine Strossen, (1995) head of the ACLU points out in her well-documented book, restrictions on free speech originally meant to protect minorities often end up being used against them.

In reviewing legal cases involving SEF in denial of tenure, it is clear that SEF can be used as an instrument of covert discrimination against minorities. As a legal handbook (Weeks, 1996) for college administrators points out, "Student ratings are an effective tool to respond to litigation involving discrimination, since they provide both qualitative and quantitative data on teaching." The handbook was suggesting that SEF can be used to show that the institution was not discriminating on the basis of gender or race but instead on the basis of SEF. In fact courts have taken the position that even if discrimination can be shown, if it can be shown that the faculty member is "really" being fired because of poor student evaluation of their teaching, discrimination is not relevant. While prime facie this is a reasonable judgement, disentangling gender and race bias from student evaluations is extremely problematic given the way evaluation forms are typically constructed, analyzed, and applied.

In conclusion, what this article suggests is that SEF is far from the benign instrument it may once have been in a more homogeneous political, gender, racial, and academically prepared student environment. Unfortunately, on many campuses the traditional model of student and teacher belongs to a past age.[40] Faculty now teach in a litigious context. The new role and impact of SEF need to be reassessed accordingly.

Summary

Validity: As currently used (and perhaps under more stringent conditions as well) the validity of SEF as used in salary, promotion, and tenure decisions, are in question relative to methodology: (1) untrained interpreters using (2) intuitive, and (3) eyeballing methods of analysis to analyze SEF, (4) not controlling for contaminating variables such as (a) level of course, (b) instructors' standards (c) grading practices, (d) subject matter or discipline, (e) personality and (f) interest of student, (g) academic level of student, (h) required course v.s elective course, (i) class size, (j) age, and (k) gender of instructor, and (l) a host of other variables; (5) Validity is also in question regarding student ability to validly render judgements regarding instruction and curricular.

Administrative Pressures To Relinquish Faculty Control: (6) SEF is typically required by administrative policy, (7) are therefore involuntary, (8) are used for salary, promotion, and tenure decisions, which allow (8) administrative intrusion into the classroom, (9) create economic incentives for shaping faculty behavior to assure (10) assent to general consumer demands for the type and level of education, (11) assuring classroom changes in accordance with educational fads, (12) political ideology, (13) to teach in a manner that results in student satisfaction, (14) leading to the retention of students, (15) their tuition dollars, resulting in (16) institutional growth.

Control Over Academic Standards: Given the above, pressures are created (17) for faculty to assent and adjust their teaching to whatever level of student the (open enrollment) institution elects to admit, (18) for conforming to student demands in the classroom by lowering standards which include (19) to lower curricular changes, (20) make tests less rigorous, (21) inflate grading, (22) fewer classroom requirements and prerequisites, (23) for easier and less course content, (24) adjusting courses to popular culture belief and notions, (25) sets up a conflict of interest between the instructor and quality of education, all of which result (26) in the opposite of the original intent of SEF which was the improvement of instruction.

General Legal Implications: Depending on their use, SEF (27) is often in conflict with usual personnel practices and procedures, (28) when published can lead to defamation of faculty reputation, (29) be discriminatory with regard to age, gender, race, and other variables, (30) involuntarily imposed, (31) not meeting what is considered contractual informed consent of faculty for their use,

Academic Freedom and Tenure: SEF can be used to (32) inappropriately dismiss competent faculty, (33) abrogate tenure, and (34) abridge academic freedom.

References

Abrami, P. C., Dickens, W. J., Perry, R. P., & Leventhal, L. (1980). Do teacher standards for assigning grades affect student evaluations of instruction? *Journal of Educational Psychology*, 72, 107-118.

Avi-Itzhak, T., and Lya K. (1986). An investigation into the relationship between university faculty attitudes toward student rating and organizational and background factors. *Educational Research Quarterly*, 10 . 31-38

Benson, D. E. Lewis, J. M. (1994). Students' evaluation of teaching and accountability: implications from the Boyer and the ASA reports. *Teaching Sociology*, 22, 195-99

Blum, D. E. (1990, October 3). U. of Guam removes evaluations from files. *The Chronicle of Higher Education* Section: Personal & Professional, p. A21

Bonetti, S. (1994). On the use of student questionnaires. *Higher Education Review*, 26, 57-64

Bures, A. L., DeRidder J. and Hsin-Min Tong. (1990). An empirical study of accounting faculty evaluation systems. *The Accounting Educators' Journal*. (Summer). 68-76.

Cahn, S. (1987, October 14). Faculty members should be evaluated by their peers, not by their students. *Chronicle of Higher Education*, p. B2

Carey, G. W. (1993). Thoughts on the lesser evil: student evaluations. *Perspectives on Political Science*, 22, 17-20

Cashin, W. E. (1996). Developing an effective faculty evaluation system. Idea Paper No. 33, Manhattan: Kansas State University, Center For Faculty Evaluation and Development (January)

Cashin, W. E. (1995). Student ratings of teaching: The research revisited. IDEA Paper No. 32. Manhattan, KS: Kansas State University, Center for Faculty Evaluation and Development.

Chacko, T. I. (1983). Student ratings of instruction: A function of grading standards. *Educational Research Quarterly*, 8(2), 19-25.

Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research*, 51, 281-309.

Copeland, J.D. and Murry, J. W., Jr. (1996). Getting tossed from the ivory tower. *Missouri Law Review*, 61, 233-327

Cohen, P.A. (1983). Comment on "a selective review of the validity of student ratings of teaching." *Journal of Higher Education*, 54, 448-458.

Cotter, William. (1996). Why tenure works. *Academe*. (Jan/Feb) 26-29

Crumbley, L.. (1996). Due process/defamation issues [Not to be construed as legal advice] Society for A Return to Academic Standards. [Online] Available: http://tamu.edu:8000/~crumble/sfrtas.html

Crumbley, L. (1995). On the dysfunctional atmosphere of higher education: games professors play. Accounting Perspectives, 1, (Spring). [Online] Available: http://tamu.edu:8000/~crumble/sfrtas.html

Crumbley, L. D. and Fliedner, Eugene (1995). Accounting administrators' perceptions of student evaluation of teaching (set) information. Manuscript, Department of Accounting, Texas A&M University.

Damron, J. C. (1996). Instructor personality and the politics of the classroom. Manuscript, Douglas College, New Westminster, British Columbia, Canada V3L 5B2. (An Earlier Versions in the June 1994 Issue of Faculty Matters (No. 5, Pages 9-12) and the September, 1994 Issue of Update (The Newsletter of the Okanagan University College Faculty Association). Douglas College)

Dawes, R. M., Faust, D., and Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, 243, 1668-1674

Dershowitz, A. (1994). *Contrary to popular opinion*. New York: Berkley Books

Dilts, D. A., Samavati, H., Moghadam, M.R., and Haber, L.J. (1994). Student evaluation of instruction: Objective evidence and decision making. *Journal of individual employment rights*, 2,73- 86.

Divoky, J. J. and Rothermel, A. (1988). Student perception of the relative importance of dimensions of teaching performance across type of class. *Educational Research Quarterly*, 12. 40-45

Dowell, D. A., & Neal, J. A. (1982). A selective view of the validity of student ratings of teaching. *Journal of Higher Education*, 53, 51-62.

DuCette, J. and Kenney, J. (1982). Do grading standards affect student evaluations of teaching? Some new evidence on an old question. *Journal of Educational Psychology*, 74. 308-314.

Education Employment Law News (1994). How big a role should student evaluations play in the assessment of a professor for tenure? (January) pp. 3-4.

EEOC V. Franklin & Marshall College, No. 84-1739, United states court of appeals for the third circuit, 775 F.2d 110; 3 Fed. R. Serv. 3d (Callaghan) 282; 85 A.l.r. Fed. 669; 39 Fair Empl. Prac. Cas. (Bna) 211; 38 Empl. Prac. Dec. (Cch) P35,644, August 5, 1985, Argued, October 21, 1985, Decided

Evaluating Teacher Evaluations (1996). Article in the Topical Interest Group: Assessment in Higher Education submitted by glen9579@novell@uidaho.edu on 11/12/96. Available: http://marsquadra.tamu.edu/TIG/GeneralArticles/evaluatingteacherevaluations.html

Faust, D., Guilmette, T. J., Hart, K., Arkes, H.R., Fishburne, F. J. and Davey, L. (1988). Neuropsychologists' training, experience, and judgement accuracy. *Archives of Clinical Neuropsychology*, 3, 145-163.

Feldman, K. A. (in press). Identifying exemplary teachers and teaching: Evidence form student ratings. In R. P. Perry & J. C. Smart (Eds.), *Effective teaching in higher education: Research and practice*. New York: Agathon Press.

Feldman, K.A. (1993). College students' view of male and female college teachers: Part II---Evidence from students' evaluations of their classroom teachers. *Research In Higher Education*, 34, No. 2. 151-191.

Feldman, K.A. (1983). Seniority and experience of college teachers as related to evaluations they receive from students. *Research in Higher Education*, 18, 3-124.

Franklin, J., & Theall, M. (1990). Communicating student ratings to decision makers: Design for good practice. In Theall, M. & Franklin J. (Eds.), *Student Ratings of Instruction: Issues For Improving Practice*. San Francisco: Jossey-Bass.

Garb, H. N. (1989). Clinical judgment, clinical training, and professional experience. *Psychological Bulletin*, 105, 387-392.

Gilbaugh, John W. (1982). Renner substantiated. *Phi Delta Kappan*, 63 (6 Feburary) 418.

Gillmore, G. (1984) Student ratings as a factor in faculty employment decisions and periodic review. *Journal of College and University Law*, 10, 557-576 .

Goldberg, G. and J. Callahan. (1991). Objectivity of students' evaluations of instructors. *Journal of Education for Business*, 66, (July/August) 377-378.

Goldman, L. (1993). On the erosion of education and the eroding foundations of teacher education. *Teacher Education Quarterly*, 20, 57-64.

Goldman, L. (1985). The betrayal of the gatekeepers: grade inflation. *Journal of General Education*, 37, 97-121.

Greenwald, A. G. (1996). Applying social psychology to reveal a major (but correctable) flaw in student evaluations of teaching. University of Washington, Draft Manuscript, March 1, 1996.

Greenwald, A.& G. Gillmore. (1996). No pain, no gain? The importance of measuring course workload in student ratings of instructions. University of Washington. Manuscript submitted for publication.

Guernsey, L. (1996, September 20) Metropolitan State College reviews its tenure policy. *The Chronicle of Higher Education*, p. A13.

Hayes, S. C. (1991). The emperor's clothes: Examining the "delusions" of professional psychology: The healthy skepticism of David Faust. *Science*, 1, 22-25.

Healy, P. (1996, January 26). Newly appointed trustees stir up public colleges. *The Chronicle of Higher Education*, p. A26.

Heller, S. (1986, December 17). Proper use of student evaluations at issue in professor's legal appeal to regain job. *The Chronicle of Higher Education*, p. 14.

Holmes, D. S. (1972). Effects of grades and disconfirmed grade expectancies on students' evaluations of their instructor. *Journal of Educational Psychology*, 63, 130-133.

Howard, G. S., & Maxwell, S. E. (1982). Do grades contaminate student evaluations of instruction? *Research in Higher Education*, 16, 175-188.

Howard, G. S., & Maxwell, S. E. (1980). Correlation between student satisfaction and grades: A case of mistaken causation? *Journal of Educational Psychology*, 72, 810-820.

Howard, G. S., Conway, C. G, & Maxwell, S. E. (1985). Construct validity of measures of college teaching effectiveness. *Journal of Educational Psychology*, 77, 187-196.

Husbands, C. and Fosh, P. (1993). Student's evaluation of teaching in higher education: Experiences from four European countries and some implication of the practice. *Assessment & Evaluation on Higher Education*, 18, 95-114.

Kemp, B. W. and G. S. Kuman, (1990). Student evaluations: are we using them correctly? *Journal of Education for Business*, 65, (November/December). 106-11.

Kolevzon, M. S. (1981). Grade inflation in higher education: a comparative study. *Research in Higher Education*, 15, 195-212.

Kresl, P. (1995). Grade inflation must be examined before discussed. [Online] Available: http://www.bucknell.edu/bucknellian/sp95/03-09-95/ops/4002.html (March 7).

Larkin, J., McDermott, J., Simon, and Simon, H. (1980). Expert and novice performance in solving physics problems. *Science*, 208, 1335-1342.

Leatherman, C. (1996a, September 13). Commission recommends strengthening college presidents' power. *The Chronicle of Higher Education*, p. A43.

Leatherman, Courtney (1996b, October 25). More faculty members question the value of tenure. *The Chronicle of Higher Education*, p. A12.

Magner, D. K. (1996, September 20) Minnesota regents' proposals stir controversy with faculty. *The Chronicle of Higher Education*, p. A11.

Magner, D. K. (1995, May 19). Mid-semester removal of professor roils university of montana. *The Chronicle of Higher Education*, p. A25.

Marsh, H. W., & Dunkin, M. J. (1992). Students' evaluations of university teaching: A multidimensional perspective. *Higher Education: Handbook of Theory and Research*, 8, 143-233.

Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, 76, 707-754.

Marsh, H. W. (1982). Validity of students' evaluations of college teaching: A multitrait-multimethod analysis. *Journal of Educational Psychology*, 74, 264-279.

Marsh, H. W. (1980). The influence of student, course, and instructor characteristics on evaluations of university teaching. *American Educational Research Journal*, 17, 219-237.

Matsuda, M. J. (1989). Public response to racist speech: considering the victim's story. *Michigan Law Review*, 87, 2320-2378.

McCallum, L.W. (1984). A meta-analysis of course evaluation data and its use in the tenure decision. *Research in Higher Education*, 21, 150-158.

McKeachie, W. J. (1979). Student ratings of faculty: A reprise. *Academe*, 65, 384-397.

McMurtry, J. (1991). Education and the market model. *Journal of Philosophy of Education*, 25, 209-217.

Menand, L. (1996). (Ed.), *The Future of Academic Freedom*. Chicago: University of Chicago Press.

Menand, L. (1993). The future of academic freedom. *Academe*. (May/June) 11-17.

Murray, B. (1996). Attacks on tenure lead to new standards: To contain spiraling higher-education costs and be fairer to minority faculty, the tenure system should be reformed, educators say. *American Psychological Association Monitor* (May).

Nelson, J. P. Lynch, K. A. (1984). Grade inflation, real income, simultaneity, and teaching evaluations. *Journal of Economic Education*, 15, 21-37.

Parini, J. (1995, July 14). Tenure and the loss of faculty talent. *The Chronicle of Higher Education*, Section: Opinion, POV.

Pennsylvania State University (1996). Policy On Confidentiality Of Student Records.[Online] Available: '//psuinfo/Policies-Rules for Students/Policies and Rules/Records. Also available via gopher from info.psu.edu.

Pinsker, S. (1989). Teaching in a litigious age. *Change*, 21, (July/August), 50-54.

Powell, R. W. (1977). Grades, learning, and student evaluation of instruction. *Research in Higher Education*, 7, 193-205.

Rabinowitz, J. (1993). Diagnostic reasoning and reliability: a review of the literature and a model of decision-making. *The Journal of Mind and Behavior*, 14, 297-316.

Rebell, M, A. (1990). Legal issues concerning teacher evaluation. In Millman, Jason & Darling-Hammond, Linda, (Eds.), (pp. 337-355) *The new handbook of teacher evaluation*. Beverly Hills: CA. SAGE.

Reisman, D. (1981). *On higher education: The academic enterprise in an era of rising student consumerism*. San Francisco: Jossey Bass.

Renner, Richard, R. (1981). Comparing professors: how student ratings contribute to the decline in quality of higher education. *Phi Delta Kappan*, 63 2 Oct. 128-30.

Robinson, R. K., and Fink, Ross, L. (1996). Public disclosure of teaching evaluations: privacy and liability considerations. *Journal of Education for Business*. May/June 284-287.

Ryan, J. J., J.A. Anderson, and A.B. Birchler. (1980). Student evaluations: The faculty responds. *Research in Higher Education*, 12 (December) 317-333.

Sacks, P. (1996). *Generation X Goes To College. An Eye Opening Account of Teaching in Postmodern America*. LaSalle, Il: Open Court.

Schrank, B. (1993). Academic freedom and the inclusive university. Paper presented at the Learned Societies Meeting, Carleton University, Ottawa, Ontario, 5 June 1993. [Online] Available: http://www.caut.ca/swc/afiu.html

Schrecker, E. W. (1986). *No ivory tower: Mccarthyism & the universities*. Cambridge: Oxford University Press.

Scriven, M. (1995). A unified theory approach to teacher evaluation. *Studies in Educational Evaluation*, 21, 111-129.

Scriven, M. (1993). The validity of student ratings: In teacher evaluation. Evaluation & Development Group.

Scriven, M. (1991). *Evaluation thesaurus*. (4th ed.) Newbury Park, Calif. : SAGE Publications.

Seldin, P. (1993, July 21). When students rate professors. *The Chronicle of Higher Education*. Opinion p. POV.

Seldin, P. (1984) *Changing practices in faculty evaluation*. San Francisco: Jossey-Bass.

Snyder, C. R., & Clair, M. (1976). Effects of expected and obtained grades on teacher evaluation and attribution of performance. *Journal of Educational Psychology*, 68, 75-82.

Sommer, J. W. (Ed.) (1995). *The academy in crisis: The political economy of higher education*. (pp.95-126) London: Transactions Publishers.

Stern, J., and Flynn, P. D. (1995). Students propose a course of action for grade inflation. *The Bucknellian*. (Feb, 20).[Online] Available: http://www.bucknell.edu/bucknellian/sp95/03-02-95/ops/4165.html

Stone, J. E. (1995). Inflated grades, inflated enrollment, and inflated budgets: an analysis and call for review at the state level. *Education Policy Analysis Archives*, 3, Number 11 (2075) [Online] Available: http://olam.ed.asu.edu/epaa/v3n11.html

Strossen, N. (1995). *Defending pornography: Free speech, sex and the fight for womens's rights*. New York: Anchor/Doubleday.

Tagomori, H. T. Bishop, L. A. 1995. Student evaluation of teaching: Flaws in the instruments. *Thought & Action*, 11, 63-78.

The Chronicle of Higher Education. (1995, March 31). South Carolina lawmaker would abolish tenure, p. A24.

The Chronicle of Higher Education (1994a, April 27). State calls for release of faculty evaluations, p. A4.

The Chronicle of Higher Education. (1994b, December 14). Faculty evaluations to go on line at U. of Texas at Austin, p. A21.

University of Michigan (1994). Toward a definition of tenure. Senate Advisory Committee on University Affairs Standing Subcommittee on Tenure Endorsed Unanimously by the University of Michigan Senate Assembly, December 12.[Online] Available: http://www.umich.edu/~aaupum/tenure05.htm

Vasta, R., & Sarmiento, R. F. (1979). Liberal grading improves evaluations but not performance. *Journal of Educational Psychology*, 71, 207-211.

Weeks, K. M. (1996). Managing departments: Chairpersons and the law. Nashville, Tennessee: College Legal Information, Inc.

Wilson, R. (1996, June 7) Scholars off the tenure track wonder if they'll ever get on. *The Chronicle of Higher Education*, p. A12.

Worthington, A. G., & Wong, P. T. P. (1979). Effects of earned and assigned grades on student evaluations of an instructor. *Journal of Educational Psychology*, 71, 764-775.

Young, R. D. (1993). Student evaluation of faculty: a faculty perspective. *Perspectives on Political Science*, 22, 12-16.

Yunker, B.J. and Sterner, J (1988). A survey of faculty performance evaluation in accounting. *Accounting Educators' Journal*, (Fall). 63-74.

Zirkel, P. A. (1996) *The law of teacher evaluation*. Bloomington, IN: Phi Delta Kappa Educational Foundation.

Notes:

1. Address correspondence and notification of commentary to: Robert E. Haskell, Ph.D. Professor of Psychology, Department of Social and Behavioral Sciences, University of New England, Biddeford, Me. 04005. Email haskellr@gwi.net. I would like to thank Larry Crumbly of Louisiana State University for providing me with some of the sources cited in this article. I would especially like to thank professors John Damron, of Douglas College, and John Stone, of the Andrew Jackson Institute, for sources and for reading an early draft of this manuscript.

2. The closest and only allusion to SEF is found in Chapter Four by Cass R. Sunstein. In discussing the formal and informal regulation of speech (i.e., certain viewpoints) the author notes, "the evaluation of students and colleagues cannot occur without resort to content, and it would be most surprising if viewpoint discrimination did not affect many evaluations." (p.106)

3. Since the issue of SEF validity, in terms of learning, is so central a few observations of the literature is necessary. Greenwald and Gillmore (1996) have categorized some of the significant reviews and empirical research that find in favor of validity of SEF as measures of quality of instruction, for example, Cashin (1995), Cohen (1981), Feldman (in press), Howard, Conway,

and Maxwell (1985), Howard and Maxwell (1980, 1982), Marsh (1980, 1982, 1984), Marsh and Dunkin (1992), and McKeachie (1979). Reviews and empirical critiques that are critical of the validity of SEF include, Chacko (1983), Dowell and Neal (1982), Holmes (1972), Powell (1977), Snyder and Clair (1976), Vasta and Sarmiento (1979), and Worthington and Wong (1979). Positions, suggesting cautious support for validity of SEF while at the same time expressing concerns about the adequacy of their support, include, Abrami, Dickens, Perry, & Leventhal (1980). The recent methodologically sophisticated research of Greenwald (1996), and Greenwald and Gillmore (1996) find strong evidence inconsistent with the common dismissive interpretation of the relationship between SEF and high student grades as reflecting a relationship between amount learned and student ratings.

4. Cahn (1987) suggests that student ratings do not measure the instructional effectiveness or the intellectual achievement of students. SEF measure student satisfaction, attitudes toward instructors course, student personality, and the psychosocial needs of the student. Cahn suggests, students know if instructors are likeable, not if they are knowledgeable; they know if lectures are enjoyable, not if they are reliable. In a meta-analysis Cohen (1983) concludes from his study, "While the magnitude of the average rating/achievement correlation for the thirty-three multisection courses is not overwhelming [14.4% of shared variance between ratings and the criterion], the relationship is certainly stronger and more consistent than we were led to believe..." (p. 455). Dowell & Neal (1982) conclude that "The research literature can be seen as yielding unimpressive estimates of the validity of student ratings. The literature does not support claims that the validity of student ratings is a consistent quantity across situations. Rather, the evidence suggests that the validity of student ratings is modest at best and quite variable (p. 59). "The variability in obtained validity coefficients even in studies with reasonable methodological requirements. . . lead us to suspect that the validity of student ratings is influenced by situational factors to such an extent that a meaningful, generalizable estimate of their validity does not exist. In general . . .no meaningful estimate of the validity of student ratings can be provided with confidence that is generalizable enough to be useful..." (60-61)

5. For example, studies demonstrate the following variables: Age, gender, class size, year of student, level of student, instructor style, subject matter, major or elective course, student interest in subject matter, instructor grading difficulty, anonymous v.s signed ratings, whether students are informed of their use, instructor present v.s instructor absent while completing the evaluation (see for example, Divoky and Rothermel, 1988).

6. At least one study found that 70% of students "indicated that the grade they thought they would get influenced the level at which they rated their professors" (in Goldman, 1985). The grade inflation literature shows that the typical percentage of A's and B's on many campuses is anywhere from 70-90%, and that the percentage of graduates who are graduating with cum laude and above is somewhere in the range of 50-70%.

7. If used correctly (see Copeland and Murry, 1996; Kemp and Kuman, 1990; Scriven, 1995, 1993, 1991; Seldin, 1984), SEF can be very useful instructionally, and when used in conjunction with other methodologically sound evaluation procedures and criteria, it can assist in informing an institution when a faculty does not pass muster as an effective teacher. The question, of course is: does student feedback to faculty result in improved teaching and student learning. In a review of studies, Marsh (1984) suggests that there is a small positive correlation for improved student learning of SEF feedback to faculty if used in a carefully constructed collegial consultation process. They are seldom used in this manner.

8. Bona Fides: I am a full professor, so I'm not personally concerned with SEF effect on

promotion, tenure, and performance evaluations. I typically receive a 3.4 or 5 on a 4. scale on my evaluations, and in the classroom I do a fairly good dog and pony show as well.

9. Other countries using SEF questionnaires seem not as concerned with research on their validity as is the U.S., there is apparently almost no published technical literature on validity and bias in SEF in German universities. There is however, a debate in German higher education as to whether SEF is an invasion of academic autonomy (Husbands and Fosh, 1993).

10. While faculty are entitled to freedom of discussion and inquiry in their classroom, it is a generally recognized limitation that they should not introduce controversial matter which has no relation to their subject. What constitutes "controversial" and "no relation," however, often remains an open question.

11. AAUP Committee C on College and University Teaching, Research, and Publication. It was adopted by the Council of the American Association of University Professors in June 1975 and endorsed by the Sixty-first Annual Meeting as Association policy. They also state: "An important and often overlooked element of evaluating teaching is an accurate description of a professor's teaching. Such a description should include the number and level and kinds of classes taught, the numbers of students, and out-of-class activities related to teaching. Such data should be very carefully considered both to guard against drawing unwarranted conclusions and to increase the possibilities of fairly comparing workloads and kinds of teaching, of clarifying expectations, and of identifying particulars of minimum and maximum performance. Other useful information might include evidence of the ability of a teacher to shape new courses, to reach different levels and kinds of students, to develop effective teaching strategies, and to contribute to the effectiveness of the individual's and the institution's instruction in other ways than in the classroom.

12. Kolevzon found an interesting breakdown of these responses focused upon the finding that faculty with fewer years of teaching experience at the university were more impressed with the importance of the role played by student evaluations when compared with senior faculty. p.208 Faculty with at least 6 years of teaching experience were more strongly supportive of the position that student evaluation forms were responsible for lenient grading. Avi-Itzhak and Lya found---perhaps counterintuitively---that senior faculty were the most opposed to SEF being used for salary, promotion and tenure.

13. It should go without saying, that not all students are the same. SEF vary by maturity, intellectual level, i.e., graduate student evaluations v. undergraduate (See, Divoky and Rothermel, 1988; Dilts, Samavati, Moghadam, and Haber, 1993) and therefore probably by campus and program. This may account in part for some of the wide variation in faculty attitudes toward SEF.

14. Other studies also show that inflating grades leads to higher SEF (See, DuCette and Kenney, 1982; Goldberg and Callahan, 1991; Kemp and Kuman, 1990).

15. Given that SEF serve a dual function of both promotion termination, another area of possible counterproductiveness is the inherent conflict in the necessary cooperation for faculty development purposes; SEF used for such purposes is thus in conflict with faculty self-protection. Unlike in the past, now that SEF are an increasingly important instrument at many institutions, its underlying problems become increasingly evident.

16. Last semester, says Dershowitz, "a small group of students complained about my

teaching...[subject]... from a civil liberties perspective." I responded that it was important for the students to hear a variety of perspectives about...[subject]..., just as they hear, without objection, about other crimes. I also reminded them that the majority of students who speak in class present the "politically correct" views. I told them that the answer to an offensive argument is not to censor but rather to come up with a better argument. A few days later, one of the students told me that I should expect to be "savaged" in the evaluations that each teacher receives from the students. (italics and brackets added, p.118-119)

17. I use these examples as merely illustrations of the role of SEF, making no attempt to assess validity or non validity of the examples.

18. I use this case because I have a certain degree of prima facie confidence in the instructor's view. I had a similar experience teaching a required abnormal psychology course to a group of vocational majors (Occupational Therapy). I have been teaching for twenty years, and received good student evaluations and have earned tenure twice at institutions that rely heavily on SEF. I recently experienced my first student "revolt." The students were under pressure not to get two D's. My course was the last course before they were formally admitted into the major. I received no support from administration, including the department chair. Moreover some of the vocational faculty (who incidently are not doctoral level faculty), making it known to the students, complained to administration about my course. For interesting similar personal accounts, See Peter Sacks (1996). Generation X Goes To College. An Eye Opening Account of Teaching in Postmodern America. LaSalle, Il: Open Court

19. I have contacted a number of legal scholars who have written compendia of legal cases regarding the denial of tenure and promotion. Few were able to refer me to cases involving SEF. I would like to express my appreciation to Professor William A. Kaplin, School of Law, Catholic University of America, and to Michael Rebell, of Rebell and Katzive, New York city.

20. I wish to than Dr. John Damron of Douglas College, for this reference.

21. Scriven has done considerable rigorous work on evaluation procedures, particularly on the justification inferring from ratings to conclusions about the merit of teaching on the basis of statistical correlations between ratings and student learning gains. He suggests that such inferences are invalid, unless a number of stringent conditions are met on the design, administration, and use of such rating forms.

22. Litigation resulting from invalid data in Great Britain has occurred. As Husbands and Frosh (1993) note, "As far as we know, there have been no comparable cases in European courts but, if European universities follow the American example of using student evaluations largely or exclusively for summative purposes, it is only a matter of time before there is external examination of the techniques being used, and of their suitability for the purposes for which they are intended." p.103. They conclude, "it is a sad commentary on the gullibility of some people in the face of numerical data that it required the intervention of the courts to force the discontinuation of the more gross forms of this type of interpretation" p. 103.

23. Many faculty are not aware of the extent of the confidentiality of student information. For example, (1) student scores or grades can not be posted publicly by name, social security numbers, or any other identifier that can be know by anyone but the instructor and student, (2) student papers or lab reports that have student names and grades can not be left in places that are accessible to others, (3) other students may not have revealed to other in a class, (4) faculty are not to request student information without a legitimate educational reason, (5) student grades or

other educational information may not be shared with other faculty unless the faculty has a specific legitimate reason to know, (6) libraries are apparently prohibited from revealing to instructors class reading material s/he has specifically put on reserve in the library for students to read to see if they have read the material, (7) student grades or other educational information can not even be revealed to the parent of the student (who may be paying for the student's education) without written permission of the student. There are many other restrictions as well.

24. According to Black's Law Dictionary (With Pronunciations Sixth Edition (1990). St. Paul, Minn.: West Publishing Co.) "Libel. A method of defamation expressed by print, writing, pictures, or signs. In its most general sense, any publication that is injurious to the reputation of another. A false and unprivileged publication in writing of defamatory material. Bright v. Los Angeles Unified School Dist., 51 Cal.App.3d 852, 124 Cal.Rptr. 598,694. "A maliciously written or printed publication which tends to blacken a person's reputation or to expose him to public hatred, contempt, or ridicule, or to injure him in his business or profession. Corabi v. Curtis Pub. Co., 441 Pa. 432, 273 A.2d 899,904. "Accusation in writing or printing against the character of a person which affects his reputation, in that it tends to hold him up to ridicule, contempt, shame, disgrace, or obloquy, to degrade him in the estimation of the community, to induce evil opinion of him in the minds of right thinking persons, to make him an object of reproach, to diminish his respectability or abridge his comforts, to change his position in society for the worse, to dishonor or discredit him in the estimation of the public, or his friends and acquaintances, or to deprive him of friendly intercourse in society, or cause him to be shunned or avoided, or where it is charged that one has violated his public duty as a public officer. Almost any language which upon its face has a natural tendency to injure a man's reputation, either generally or with respect to his occupation. Washer v. Bank of America Nat. Trust & Savings Ass'n, 21 Cal.2d 822, 136 P.2d 297, 300."

25. According to Rebell (1990), Florida specifically mandates the school board to provide professional training programs to "ensure that all individuals with evaluation responsibilities understand the proper use of the assessment criteria and procedures" (Fla. Educ. Code, /sec 231.29(2)). p.345. Apparently, however, such mandates are not applied to higher educational institutions.

26. Dowell & Neal (1982) suggest that "The attraction of student ratings to higher education officials lies in the seductive way they seem to reduce a complex human activity, teaching, to simple numbers. Administrative over reliance upon ratings may encourage faculty to "perform to criterion," to teach in such a way as to accrue good ratings, even though this may not result in optimal educational practice. Ratings cannot substitute adequately for more informed judgment about teaching effectiveness, which might be obtainable from reviewing course materials, visiting classes, and discussing teaching practices with faculty. Unfortunately, these detailed qualitative methods are currently regarded as too expensive and difficult to use in most institutions. Evidence presented in [Cohen's 1983] review reinforces our earlier conclusion that student ratings are inaccurate indicators of student learning and therefore are best regarded as indices of "consumer satisfaction" rather than teaching effectiveness" (p.60-61)

27. Recently, a group of college and university presidents have proposed that faculty input into decisions be reduced. The commission's 50-page report, "Renewing the Academic Presidency: Stronger Leadership for Tougher Times," makes recommendations to presidents, professors, trustees, and public officials. Copies of the report, available for $12.95 for members of A.G.B. and $19.95 for non-members, may be obtained from the Association of Governing Boards of Universities and Colleges, One Dupont Circle, Suite 400, Washington 20036; (202) 296-8400. From: Leatherman, (1996a)

28. It is perhaps ironic in a democracy that there should be so much scrutiny of faculty behavior in the classroom. One faculty who has taught and lived in totalitarian states observed (in Sacks, 1996), " I grew up under a totalitarian government and worked as a teacher there, and never did anybody come to visit my classroom to make sure I was toeing the line in that sense...there was more "oppression" of teachers the United States, a country that boasts the ideals of freedom speech and thought, and that there was more pressure to conform to "acceptable" ideas in U.S. classrooms." (p.33)

29. See endnote # 12

30. For purposes of this paper I consider the mandated use or non use of SEF for salary, promotion and tenure by union contract agreements a special case of the "voluntary" nature of SEF. Some union contracts prohibit their use in salary, promotion and tenure decisions. Other require that students sign their evaluations.

31. There is all too frequently negative administrative support for faculty when students complain. For years I've seen faculty struggle with this issue. In the end, they usually conform to it---at least to some degree. When I informed one administrator that I was going to change my grading system from my already drastically modified Bell curve to to the campus norm of giving mostly A's and B's, because I just could no longer "compete" for students in my courses, he soberly looked me straight in the eye and firmly said, "That's a good idea."

32. See, Grimes v. Eastern I11. Univ., 710 F.2d 386, 388 (7th Cir. 1983). Cited in, Copeland and Murry, (1996).

33. A recent article on the nation-wide attacks on tenure, repeatedly referred to it as "job security" for professors. It was noted that American Association for Higher Education (AAHE) researchers found that tenure serves an important purpose by "rewarding dedicated, productive professors." Other articles maintain that a tenure system helps to recruit and retain quality faculty. And a representative of the American Association of Psychology Professors, argued that "tenure is only awarded to faculty who conduct quality research and teaching" (Murray, 1996) For similar justifications see Leatherman (1996b). After talking about job security in Business and in the military, both of which have nearly vanished, one faculty said "I can't say there's some obvious reason why higher education should be immune to it"(in Leatherman,1996b). As a consequence, there seems to be a merging of the popular culture and the academic views of tenure, with the direction of flow from the popular culture to the academic. Arguments in defense of tenure increasing exhibit a tone of, "...and, oh! by the way, it's to help insure academic freedom as well." faculty could be given five and ten-year contracts. While this may address the irrelevant issue of job security in some measure, it neither solves the problem, nor addresses the issue of academic freedom. This is why it doesn't: First---and this seems to be not widely recognized---it merely postpones being "dehired." Second, it exacerbates the job security issue. By the time a faculty has been at an institution for ten years, s/he is either an associate professor, or is at a salary range that reduces the chances of finding another position. Increasingly, with institutions cutting costs, most positions are entry level ones. So variable length contracts just provide an additional cost cutting mechanism for eliminating senior ---or soon to be senior---faculty. And third, as a consequence, such contracts tend to put a damper on faculty feeling free to examine and critique issues that are unpopular. With such contracts, the tenure function does not in fact exist.

34. I am not an uncompromising academic isolationist advocating the development of a kind of

Monroe Doctrine applied to higher education. I see the value of some changes that have taken place in the university which could probably only have occurred as the result of outside pressures. But such changes should be made with extreme caution, reason, logic, and evidence. Most importantly, any changes should made within the boundaries of what a university needs to be in order to perform its role in society. To tamper with tenure is to clearly cross one of those boundaries.

35. This is not intended as a blanket apologia for academia. There are many problems within the academy. In many other areas, I am a severe critic of my colleague's collective behavior.

36. Increasingly, and perhaps understandably, junior faculty are questioning the system of tenure. After all, not only do they not have it, but it is becoming increasingly difficult to attain, politically and in many other ways. A recent survey by the Higher Education Research Institute at the University of California at Los Angeles shows that more than a third of some 34,000 faculty either strongly or somewhat agreed that tenure is an outmoded concept (an increase from 1989), seeing it as "elitist," a "holdover," and a tiered system of "haves and have nots" (in Leatherman, 1996b). But it seems that even older faculty are beginning to question tenure. The foundations of such reasoning, it seems to me, is very egocentric, shortsighted, and dangerous to the very concept of a university in a democratic society.

37. McMurtry delineates the inappropriate use of the consumer metaphor in education. "The best product on the market, as we know, is the one which is the most 'problem-free' for its purchaser--delivered ready made for instant easy use','guaranteed replacement' if it does not work, and 'repaired cost-free' whenever it needs maintenance attention. The best education, on the other hand, is the opposite on all standards of excellence. It cannot be produced or delivered by another at all, is never ready-made nor instant, and cannot be guaranteed replacement or service cost-free if it is not working. The higher the standards it has, the less it can be immediate in yield, the more work it demands of its owner, and the more its failures must be overcome by its possessor's own work. An education can never be 'problem free', and poses ever deeper and wider problems the higher the level of excellence it achieves. Freedom in the market is the enjoyment of whatever one is able to buy from others with no questions asked, and profit from whatever one is able to sell to others with no requirement to answer to anyone else. Freedom in the place of education, on the other hand, is precisely the freedom to question, and to seek answers, whether it offends people's self-gratification or not....What is the best policy for buying a product--to assert the customer's claim ‘as always right'--is the worst possible policy for a learner. What is the best policy for selling a product--to offend no-one and no vested interest--ma): be the worst possible policy for an educator. The principles of freedom here are contradictory, and become the more so the more each is realized." p.213-214

38. Extending the ancient Greek account, in computer science a Trojan Horse refers to a set of instructions hidden within a legitimate program, causing a computer to perform illegitimate functions.

39. A recent edited book on The Future of Academic by Louis Menand (1996) provides an extensive series of closely reasoned chapters addressing the multicultural-based attacks on the First Amendment. See especially Henry Louis Gates, Jr.'s "Critical Race Theory and Freedom of Speech."

40. See, for example Pinsker, (1989).

About the Author

Robert E. Haskell

Brief Bio ~ Robert E. Haskell has been teaching college and university level courses for over twenty years. He earned his Ph.D. from the Pennsylvania State University in Psychology and Social Relations, his M.A., and B.A. from San Francisco State University. His areas of research and teaching include: transfer of learning, analogical reasoning, small group dynamics. Major publications include: four books, the latest of which is, *The Future of Education and Transfer of Learning: A Cognitive Theory of Learning and Instruction For The 21st Century* (forthcoming), and numerous presentations, chapters, and research articles in national and international journals. He also serves on several editorial review boards, and is Associate Editor of The Journal of Mind and Behavior. He is former Chair, and currently Professor of Psychology, Department of Social and Behavioral Sciences, University of New England.

Professor of Psychology
University of New England
Biddeford, Maine 04005

UNE Home Page: http://www.une.edu/sbs/haskell.html

E-mail: rhaskell@javanet.com

Copyright 1997 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://olam.ed.asu.edu/epaa

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, Glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692)

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Marshall University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Rocky Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
Arizona State University

Dennis Sayers
University of California at Davis

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina--Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher Education

Mary P. McKeown
Arizona Board of Regents

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois--UC

Robert T. Stout
Arizona State University

---

archives | abstracts | editors | board | submit | comment | subscribe | search

EPAA Commentary

Contributed Commentary on
Volume 5 Number 6: Haskell *Academic Freedom, Tenure, and Student Evaluation of Faculty*

and

Volume 5 Number 8: Stake *Response to Haskell: Academic Freedom, Tenure, and Student Evaluation of Faculty*

---

21 May 1997

Michael Theall

On drawing reasonable conclusions about student ratings of instruction: a reply to Haskell and to Stake. .

Kenneth Ebel was right when he said (1983, p. 65) "No corner of the university.lacks faculty members who fulminate against student evaluations, with little or no examination of the large body of research ...that underlies the practice." Lately, there has been a lot of discussion attempting to causally like student ratings to problems such as "grade inflation," and to suggest that ratings somehow violate the great traditions of academic life. I do not mean to trivialize these issues. They are important and deserve discussion. Rather, I am targeting the use (perhaps, misuse) of these issues in efforts to discredit or do away with faculty evaluation practices in general and student ratings in particular.

"Grade-inflation" (i.e., the rise in "average" grades since 1990 or is it 1950, or 1890, or the dawn of time) is always a hot topic because it's easy to make sweeping generalizations about "today's students," especially in comparison to some more-than-likely-mythical group of "better" students from some time past (especially a time "...when I was a student"). While it is probably true that a lot of things have changed about higher education including the standards used to grade, the expectations of both faculty and students, and the overall grade profile, it is very risky to point to one reason for these changes. There just isn't evidence for such a simplistic leap.

When coupled with the sometimes hysterical rhetoric about student evaluations, the inflation complaint itself gets inflated. "Ratings" become the cause of the downfall of higher education. They are, as one anonymous and angry professor wrote to me "...a corrupt practice of the 60's....A sop to students from administrators who are unwilling or unable to do anything to really improve teaching." More recently, in an article ripe with rhetoric and loaded with misinterpretations of the literature, Robert Haskell (1997) suggested that ratings are a violation of academic freedom (and of course, the cause of grade inflation). Haskell included a long quote from a chapter Jennifer Franklin and I (1990) did in our New Directions for Teaching and Learning issue #43 (Theall & Franklin, 1990b) While our point was that ratings practice must be improved, Haskell used the quote to supplement other citations as evidence in building his case that ratings were an essentially unreliable form of data that should be done away with ... a gross misinterpretation of our intent.

Frankly, I don't like to recommend articles like this to those not actively involved in ratings research or practice because such writings can mislead readers who aren't really familiar with the cited ratings literature. I'm sorry if that sounds elitist: it isn't intended to be, but I do have a reason for noting it. We (Franklin & Theall, 1989) found that ignorance of evaluation/measurement literature and methods correlated significantly with negative faculty opinions about students and student ratings. I note this because discussions about ratings are so often filled with misinformation. For example, Haskell, (in reference note #26) says that "...Cohen's (1983) review reinforces our earlier conclusion that student ratings are inaccurate indicators of student learning...". Peter Cohen's most recognized contribution to the ratings literature is the meta-analysis of multisection validity studies in which performance on a common final exam was correlated with ratings (1981). There was a .43 correlation between ratings and performance on those exams. In other words, sections with highly rated instructors had higher average scores which could not be attributed to differential grading standards or sampling errors. This is evidence for the validity of ratings, the exact opposite of what Haskell says. Calling that relationship grade inflation is simply ridiculous!

Academic freedom has been defined in many ways, but never before in a way that suggests the construct (tradition? principle? tenet?) is vulnerable to the influence of student ratings. The only comparable complaint I've heard came from a faculty member who missed classes frequently because he attended numerous religious services. He told me that since the ratings instrument contained an item about the instructor being available and his ratings on that item were depressed due to his frequent absences, the student ratings process violated his constitutionally guaranteed freedom of religion. Fortunately, in a reply to Haskell in a later issue of EPAA (V5#8), Jeffrey Stake (1997) does us the favor of providing a strong counter-argument to the academic freedom issue. Unfortunately, he then proceeds to propose that "Almost anything that can be done to undermine the administrative practice of getting students to evaluate teaching ought to be done" (p. 1). Stake says that ratings undermine the faith and trust students must place in teachers. Given his opinion of ratings, I assume that he must then feel that teachers have no need to place faith or trust in their students. This paternalism is antiquated and unrealistic. Stake also proposes that administrative use of ratings creates an image that institutions are asking if they "guessed right" about the faculty they hired. Drawing this in the other direction, I suppose that assigning grades creates the image that institutions are asking if they "guessed right" about the students they admitted. Finally, Stake proposes that asking students' opinions "... focuses the attention of students on the acting and special effects, rather than the message." (p. 2) I haven't seen many evaluation forms which contain items on either of these topics. Stake's reference is probably to "educational seduction," the skill of the infamous "Dr. Fox" who supposedly entertained students and received high ratings despite the fact that he delivered no content (Naftulin, Ware, & Donnelly, 1973). Many who do not care for ratings find one study that supports their position but ignore subsequent work (e.g. Perry et. al., 1979; Marsh & Ware, 1982) which points out problems with the original study and proceeds to clarify the issue. In this case, the determination that while style is no substitute for substance, "expressiveness" in a lecture can be a powerful aide to students' motivation and attention during lectures and thus, to recall and other aspects of eventual performance.

Jennifer Franklin and I did two studies which involved looking at grades and ratings: one specifically on grade inflation (1991) and one on the attributions of students about their academic performance (1990a). In the first study we looked at thousands of courses from one institution over a five year period during which mandatory ratings were initiated. We found a statistically significant but numerically minute increase in average grades but we also found that average ratings actually decreased in the same time period. So, even if faculty were trying to "buy" higher ratings by giving higher grades, students "weren't buyin' it."

In the attribution study we found that ratings of instructors were very consistent (i.e. reliable) across all grades. In other words, the ratings were essentially the same from 'A' students and 'F' students. This held up when class average ratings showed that the instructor was considered to be "among the best" as well as when the instructor was considered to be "average" or "among the worst." There was no revenge for low grades or reward for high grades.

Major reviews (e.g., Marsh, 1987) have reported either near zero correlations between workload, difficulty, and ratings, or positive relationships indicating that more work and/or more difficult courses often receive higher ratings (e.g., Cashin, 1988). Greenwald & Gillmore (1996a, 1996b) however, report a positive grades-ratings relationship (i.e., higher grades - higher ratings) and a negative grades-workload relationship (i.e., higher grades - less work). However, they do not then propose a direct, negative ratings-to-workload relationship (i.e., less work- higher ratings) although they suggest it when they say (p. 14) "...finding a negative path between expected grade and workload is critically diagnostic of a causal effect of grading leniency." One might extrapolate that if less/easier work is related to higher grades, and higher grades are related to higher ratings, then less/easier work should also relate to higher ratings. Whether the syllogism holds remains a question. Even if it does, there is still a question about its implications for the student ratings process. Does a relationship - even a causal one, mean that ratings are inherently invalid or unusable? Greenwald and Gillmore note that ratings are imperfect but that flaws are "correctable" (1996b) and suggest making improvements rather than dismissing ratings. For the record, the relationship between grades and ratings in other studies is relatively consistent, with correlations normally between .3 and .4 (e.g., Feldman, 1976).

The counter-argument to the grades-ratings-inflation idea is that there is (and should be) a relationship between grades and ratings because good teaching begets good learning. Good learning results in both good grades and satisfied learners. This is still a compelling notion but Greenwald and Gillmore (1996A, 1996B) have clarified the definitions and controlled some of the variables more precisely. On the surface, their work suggests that there is a relationship which does not rely on the counter-argument and that when all else is held constant, giving higher grades will result in getting higher ratings. In other words, you can't account for the relationship solely on the basis of the "good teaching begets good learning" counter-argument.

One question about their findings is whether their finding is a psychometric, an instructional, or a psychological phenomenon. Consider that ratings are undeniably a measure of the satisfaction of learners with their learning experience perhaps more than they are a direct or absolute measure of the total quality of instruction. Though this sounds like heresy from a ratings proponent, it is a position that most researchers have held for some time. Consider also, that ratings are collected before students get their final grades and thus, their opinions must be based on expectations which are, in turn, based on performance to date. Ratings thus can not be said to reflect a disconfirmed expectancy about the overall course outcome (i.e., the course grade). This would only be the case if ratings were gathered after final grades were distributed and the final grade disconfirmed what was expected as a result of the experiences during the semester. So the ratings relationship is limited to experiences and results during the term rather than to the final grade.

So what could account for the grades-ratings relationship if not "better teaching begets better learning"? I propose the following.

Given that student satisfaction is related to the students' perceptions that they have received something of value in return for their tuition dollars, and given that a series of successful classroom and related experiences has unfolded during the semester, it is reasonable to assume that students will believe that they have learned something, will be satisfied, and will provide positive ratings. Now, in many classes (especially the lower level, undergraduate courses that populate many evaluation databases and by virtue of their enrollments and numbers, may disproportionately influence results) students are not the best judges of the breadth and depth of the instructor's knowledge or the extent to which the instructor has provided a complete, current, or even adequate treatment of the subject. If students have received good grades and positive feedback from their instructors they should: 1) be satisfied; 2) feel that they have learned something; and 3) be able to honestly rate their experiences and their instructors highly [even if the experience might be marginal using some other rating criterion coming from some other group]. The data they provide is thus both valid and reliable. Valid, because they are the appropriate providers of data and (we will assume for the moment) the instruments used ask at least face valid questions. Reliable because there will be little variance across responses (Marsh, 1987) and because their responses will not change much over time (Frey, 1976).

The question now becomes one which deals with: 1) the appropriateness of course content; 2) the standards used for grading; and 3) the question of whether lenient coverage and grading were deliberately chosen in order to influence ratings. The first two items can be and should be dealt with via curricular mechanisms such as departmental review of courses and content, and faculty agreement on standards for student work. The third item represents an ethical dimension that has much less to do with ratings than it does with behavior. If it is possible to manipulate students covertly (as in "dumbing down" courses so that students are led to believe that they are learning a lot when, in fact, they are only scratching the surface), then the problem belongs to the instructor and/or the department. Don't blame students or the ratings process. If the situation involves overt manipulation (as in making a 'ratings for grades' deal), then the problem belongs to the instructor, the students, the institution, and to higher education itself. Blame the perpetrators and those who let them get away with it, not the process of collecting, reporting, and using ratings data. Ultimately, I believe this suggests that ratings are reliable and valid indicators. If what they indicate doesn't please those who get and use the data, then the users, their departments, or their institutions should correct the situation, rather than discard the data or the processes of collecting, analyzing, or reporting it.

There are three points here. The first is that no system, however well constructed is safe from manipulation. The second is that there are probably legitimate instances of situations in which the established findings from the ratings literature are violated even though these situations are infrequent enough to influence the results of large database analyses which have been the bases of most research. The third is that ratings are a symptom, not a cause. Many opinions about ratings (like those of Haskell and Stake) seek to "blame" ratings for phenomena like grade inflation. Legitimate concern for the quality of teaching and learning should prompt us to examine the contexts of teaching and learning as well as the ratings that are provided by students, and to carefully review the other available kinds of data before jumping to erroneous conclusions about what "causes" what.

Robert Haskell suggests that we do away with ratings. Perhaps the weakest part of his article is what isn't there: constructive suggestions for improvement. The thrust is simply that we should do away with student ratings. One could just as well argue that we should do away with grades. Either way, no more nasty inflation! WRONG! There are appropriate reasons and uses for both. What's needed is more substantial research, less rhetoric, and better informed faculty and administrators who can distinguish between opinion and evidence, separate mythology from established findings, and explore the situational context in order to arrive at fair and equitable conclusions about faculty performance. Doing this won't violate anyone's rights. In fact, it should lead to better evaluation, and if data are carefully collected, reported, and used, better teaching as well.

REFERENCES

Cashin, W. E. (1988) Student ratings of teaching: a summary of research. (IDEA paper No. 20) Kansas State university, Division of Continuing Education. ERIC # ED 302 567

Cohen, P. A. (1983) Comment on a selective review of the validity of student ratings of teaching. Journal of higher Education, 54, 448-458.

Cohen, P.A. (1981) Student ratings of instruction and student achievement: a meta-analysis of multisection validity studies. Review of Educational research, 51, 281-309.

Ebel, K. (1983) The aims of college teaching. San Francisco: Jossey Bass.

Feldman, K. A. (1976) Grades and college students' evaluations of their courses and teachers. Research in higher Education, 4, 69-111.

Franklin, J. L. & Theall, M. (1991) Grade inflation and student ratings: a closer look. Paper presented at the 72nd annual meeting of the American Educational Research Association. Chicago: April 7. ERIC # ED 349 318

Franklin, J. L. & Theall, M. (1990) Communicating student ratings to decision makers: design for good practice. In M. Theall & J. Franklin (Eds.) Student ratings of instruction: issues for improving practice. New Directions for Teaching and Learning # 43. San Francisco: Jossey Bass.

Franklin, J. L. & Theall, M. (1989) Who reads ratings. Knowledge, attitudes, and practices of users of student ratings of instruction. Paper presented at the 70th annual meeting of the American Educational Research Association. San Francisco: March 31.

Frey, P. (1976) Validity of student instructional ratings. Does timing matter? Journal of Higher Education, 47 (3), 327-336.

Greenwald, A. G. & Gillmore, G. M. (1996a) No pain, no gain? The importance of measuring course workload in student ratings of instruction. Journal of Educational Psychology (in press). Available HTTP: http://weber.u.washington.edu/~agg/paingain.htm

Greenwald, A. G. & Gillmore, G. M. (1996b) Applying social psychology to reveal a major (but correctable) flaw in student evaluations of teaching Available HTTP: http://weber.u.washington.edu/~agg/asptramf.htm

Haskell, R. E. (1997) Academic freedom, tenure, and student evaluations of faculty: galloping polls in the 21st century. Educational Policy Analysis Archives V5 #6. Available HTTP: http://olam.ed.asu.edu/epaa/v5n6.html

Marsh, H. A. (1987) Students' evaluations of university teaching: research findings, methodological issues, and directions for future research. International Journal of Educational Research, 11, 253-388.

Marsh, H. W. & Ware, J. E. (1982) Effects of expressiveness, content coverage, and incentive on multidimensional student ratings scales. New interpretations of the Dr. Fox effect. Journal of Educational Psychology, 74, 126-134.

Natuflin, D. H. Ware, J. E., & Donnelly, F. A. (1973) The Dr. Fox lecture. A paradigm of educational seduction. Journal of Medical Education, 48, 630-635.

Perry, R. R., Abrami, P. A., & Leventhal, L. (1979) Educational seduction: the effect of instructor expressiveness and lecture content on student ratings and achievement. Journal of Educational Psychology, 71, 107-116.

Stake, J. E. (1997) Response to Haskell: academic freedom, tenure, and student evaluations of faculty. Educational Policy Analysis Archives V5 # 8. Available HTTP: http://olam.ed.asu.edu/epaa/v5n8.html

Theall, M. & Franklin, J. L. (1990a) Attributions or retributions: student ratings and the perceived causes of performance. Paper presented at the 71st annual meeting of the American Educational Research Association. Boston: April 16. ERIC #ED 319 764

Theall, M. & Franklin, J.L. (1990b) Student ratings of instruction: issues for improving practice. New Directions for Teaching and Learning # 43. San Francisco: Jossey Bass.

About the Author

Michael Theall, Ph.D.
Associate Professor & Director
Center for Teaching and Learning University of Illinois At Springfield
Springfield, IL 62794-9243

phone: 217-786-7157
fax: 217-786-7188
email: theall@uis.edu

Education

- BA St Anslem College (1963; English)
- MS Syracuse Univ. (1976; Instructional Technology)
- PhD Syracuse Univ. (1980; Instructional Design, Development, & Evaluation)

EPAA Commentary

Contributed Commentary on
Volume 5 Number 6: Haskell *Academic Freedom, Tenure, and Student Evaluation of Faculty*
and
Volume 5 Number 8: Stake *Response to Haskell's "Academic Freedom ... & Student Evaluation"*

On Michael Theall's (and implied et al.) "A Reply to Haskell and to Stake"

Robert E. Haskell

In "On Drawing Reasonable Conclusions about Student Ratings of Instruction: a Reply to Haskell and to Stake," Michael Theall (1997), Director of the Center for Teaching and Learning at the University of Illinois at Springfield, opens his critique of my article (Haskell, 1997a) by lamentably describing my piece as rhetorical. He then goes on to characterize my article: as an example of (1) faculty who "fulminate" against SEF , as (2) "simplistic," (3) "loaded with misinterpretations of the literature," as (4) "mythology," (5) exhibiting an "ignorance of evaluation/measurement literature," as containing (6) "sweeping generalizations," (7) "misinformation," as (8) "simply ridiculous!" (9)" ripe" with "hysterical rhetoric," as (10) assuming a "mythical group of better students" of some bygone era, asserting that (11) SEF "are the cause of grade inflation," as suggesting (12) "we do away with ratings." Continuing, Theall wrote that (13) "Perhaps the weakest part of his article is what isn't there: constructive suggestions for improvement," that I (14) "suggested that ratings are a violation of academic freedom," and finally---but not exhaustively---(15) that "Academic freedom has been defined in many ways, but never before in a way that suggests the construct (tradition? principle? tenet?) is vulnerable to the influence of student ratings." The last two items, at least, are correct. With this said, let me say I am perplexed by Theall's response.

I am perplexed for three reasons. First, because of what my article actually clearly said and did not say, second, because of Theall's own findings, and third because of the tone. Let me begin with what my article actually clearly said and did not say, and Theall's reaction to it. This will lead naturally into his own findings which, ironically, seem to support much of my thesis.

Theall's Inaccuracies and Misinterpretations

In anticipation of responses like Theall's, I clearly laid out the boundaries of my article quite carefully. Evidently, he either did not notice these boundaries or for some reason elected to ignore them. I will start with the more general accusations and proceed to the more specific, both of which , as I will document, are filled with inaccuracies. He says of my article, among other things that "Perhaps the weakest part of his article is what isn't there: constructive suggestions for improvement." I had indicated that I would not be addressing this issue, nor dealing in detail with findings on validity of SEF, or with grade inflation. I clearly indicated the boundaries: I said

> In explicating SEF , many closely related issues must be substantially bracketed. These related issues include (1) its validity (Cahn , 1987; Damron , 1996; Greenwald, 1996; Greenwald and Gillmore , 1996; Scriven , 1993; Seldin , 1984; Tagomori , Bishop, and Laurence. 1995), (2) the problem of defining teaching effectiveness, (3) general

> variables affecting SEF scores, (4) alternatives to SEF 's (5) classroom politically correct or popular standards and perceptions, (6) low student academic preparation, (7) age and gender discrimination issues (Feldman, 1983, 1993) (8) strategies for change, and (9) other integrally related issues such as their being largely responsible for lowered course standards, and grade inflation. Though these are important and related issues, they can only be addressed here in so far as they directly impact the focus on SEF and academic freedom.

Despite these clearly stated boundaries, and despite the fact of providing example citations to many of the areas like validity that I could not cover---I refer the reader to endnotes #3, 4, 5, 6 and 14 in my paper. Disregarding these limits, Theall proceeded to critique my article by concentrating on (a) validity and (b) grade inflation, exactly those areas that I said would have to be bracketed. I refused to be pulled into these two areas in my paper, and I refuse to be hauled into them now.

Both validity and grade inflation are too important to be briefly dealt with within the confines of this response by simply offering, as did Theall, a citation here and a citation there. Such an ad hoc approach to these important issues is little more than a statistical analogue of using verbal anecdotes to resolve the issue. In a more appropriate forum, if Theall would like to come up with something a little more systematic than the statistical anecdotes he offered, it might be worth responding. (Note 1) In any event, Theall refused to substantially address the main issue of my article: academic freedom and SEF .

I will now address some of the undocumented specific inaccuracies of Theall's critique. From numerous places in his response, he says,

> Robert Haskell suggests that we do away with ratings....The thrust is simply that we should do away with student ratings.... I am targeting the use (perhaps, misuse) of these issues in efforts to discredit or do away with faculty evaluation practices in general and student ratings in particular....building his case that ratings were an essentially unreliable form of data that should be done away with.

I did not say that SEF should be done away with. In fact in (a) my original article and (b) my response to Stake, (Haskell, 1997b) both of which Theall evidently read, I clearly stated:

> It is important to note at the outset, that it is not SEF per se that is the issue, but the impact of its use on salary, promotion, tenure decisions, and its impact on the delivery of quality education.

In addition in endnote #6 I said, again:

> If used correctly (see Copeland and Murry, 1996; Kemp and Kuman, 1990; Scriven, 1995, 1993, 1991; Seldin, 1984), SEF can be very useful instructionally, and when used in conjunction with other methodologically sound evaluation procedures and criteria, it can assist in informing an institution when a faculty does not pass muster as an effective teacher.

Let me make it clear once again: I am not against SEF , only their use for administrative purposes. As I said in my paper, while it is true that

> SEF may indeed validly describe incompetence...given (a) the conflicting data on their validity, (b) the way many institutions have constructed SEF instruments, (c) the often unsystematic statistical method by which SEF are interpreted, and especially (d) given

> the considerable weight accorded negative comments by only a few students in making tenure and promotion decisions, it would seem SEF can all to easily be used as a covert instrument for the elimination of tenure candidates and other faculty who may threaten student tuition dollars and perhaps ideological and popular culture agendas.

Note that I said "conflicting data on their validity," not that the literature demonstrates conclusively that they are completely invalid.

Given the number of instances on which Theall used the word "practical," I would think, given the time and resource intensive requirements for rendering SEF use valid, he would appreciate the fact that the likelihood of its appropriate use is not high. (In fact he does. See below). Theall , further claims that I make causal attributions from SEF to grade inflation. He wrote,

> Lately, there has been a lot of discussion attempting to causally link student ratings to problems such as grade inflation'.... Ratings" become the cause of the downfall of higher education.... Robert Haskell (1997) suggested that ratings are.... the cause of grade inflation)....it is very risky to point to one reason for these changes.

Nowhere in my article was it either suggested that SEF is "the cause" or was the term "cause" used or such an attribution made about SEF. I did suggest that SEF is

> largely responsible for lowered course standards, and grade inflation....are responsible for a considerable amount of grade inflation.

As scholars and rigorous researchers will likely agree there is a considerable difference between attributing a single causal connection and suggesting they contribute to grade inflation. I guess the statistical concept of variance was conveniently forgotten.

Theall claims that I "blame" students. He wrote, "Don't blame students." In anticipating such a reading of my article, I stated in the opening that to question SEF is often

> seen as not only a novel idea, but as an attack on either students, or a general attack on evaluating faculty.

I evidently anticipated this claim correctly. In further anticipation of such a reading of my article, in endnote #12 I said,

> It should go without saying, that not all students are the same. SEF vary by maturity, intellectual level, i.e., graduate student evaluations v. undergraduate.

Then later in the text I stated,

> Many students understand the above described ensuing consequences. A glance at articles from online student newspapers reveals strong sentiments against what many students consider the erosion of standards created by SEF . One student writer went so far as to say "We therefore suggest a boycott of the 1995 student/teacher evaluations. This boycott will provide a more effective means of communication than anything written on the evaluation itself. Something must be done about the trend of grade inflation. We as students refuse to contribute to the downfall of academia (Stern and Flynn, 1995) Some students are thus quite aware of the effects of SEF on their education.

Theall then goes on to claim that if there is grade inflation that one of the parties we should blame is faculty; that we should "Blame the perpetrators." There is a section in my paper entitled, "Faculty Complicity and Adaptation to SEF," in which I analyze this problem, but he does not mention any of the arguments I advance. I come to quite a different conclusion, however, than does Theall .

To my amazement, Theall simply notes, almost in passing, that I suggest that SEF "somehow violate the great traditions of academic life." What does he mean, somehow? The entire article was a fairly detailed "let me count the ways"--a listing, explaining, and accounting of just how I suggest SEF do violate academic freedom, both in the abstract and with concrete examples. So when he simply says of my article that "Academic freedom has been defined in many ways, but never before in a way that suggests the construct (tradition? principle? tenet?) is vulnerable to the influence of student ratings," I would suggest that Theall, tell us more about SEF and academic freedom, since this was the main thesis of the article.

I would now like to note very briefly and comment on some other claims made by Theall . My comments will be brief because, again, I dealt with them in my paper. First, he says that students "are the appropriate providers of data," mainly because they are paying "tuition dollars." I addressed these issues in my paper in discussing the misplaced (a) political metaphor of democracy applied to instruction, and (b) corporate metaphor of student as consumer.

Theall also claims that

> ratings are collected before students get their final grades and thus, their opinions must be based on expectations which are, in turn, based on performance to date. Ratings thus can not be said to reflect a disconfirmed expectancy about the overall course outcome (i.e., the course grade).This would only be the case if ratings were gathered after final grades were distributed and the final grade disconfirmed what was expected as a result of the experiences during the semester. So the ratings relationship is limited to experiences and results during the term rather than to the final grade.

In many if not most cases Theall's claim is incorrect. This is why: (a) In many, and maybe most, courses students receive test grades and the class curve/distribution (or at least should) throughout the semester, (b) the student grapevine is quite efficient is informing students what instructors grading practices are. So ratings thus can be said to reflect a disconfirmed expectancy about the overall course outcome.

Theall at least did make one technically valid set of points. Assuming SEF are on their face valid and reliable because there is "little variance across responses", says Theall ,

> The question now becomes one which deals with: 1) the appropriateness of course content; 2) the standards used for grading; and 3) the question of whether lenient coverage and grading were deliberately chosen in order to influence ratings. The first two items can be and should be dealt with via curricular mechanisms such as departmental review of courses and content, and faculty agreement on standards for student work.

Granting for the moment SEF validity and reliability (I probably would even really accept their reliability), I would like to address 1) and 2) in the above quote (I addressed the third above). On a collective level, I agree with Theall that faculty have not been responsible in developing and enforcing standards. In fact in endnote #34 I wrote:

> This is not intended as a blanket apologia for academia. There are many problems within the academy. In many other areas, I am a severe critic of my colleague's

collective behavior.

Theall's Work Supports Much of My Article

I noted at the beginning of this response that I was perplexed at much of Theall's reaction. I am particularly perplexed at his response to a passage I cited from his work. He says,

> Haskell included a long quote from a chapter Jennifer Franklin and I (1990) did in our New Directions for Teaching and Learning issue #43 (Theall & Franklin, 1990b) While our point was that ratings practice must be improved, Haskell used the quote to supplement other citations as evidence in building his case that ratings were an essentially unreliable form of data that should be done away with ... a gross misinterpretation of our intent.

I am perplexed for two reasons. First, it is really neither here nor there that the quote apparently does not reflect Theall's (et al.) intent. Whatever an author's intent was in using a quote does not necessarily reflect on other uses of the material, especially when the quote is not used to suggest an authors position.

Second, I am perplexed because much of Theall's chapter revolves around other material that he suggests demonstrate some of the severe problems in using SEF as they are currently used. The full quote to which Theall refers is as follows:

> Even given the inherently less than perfect nature of ratings data and the analytical inclinations of academics, the problem of unskilled users, making decisions based on invalid interpretations of ambiguous or frankly bad data, deserves attention. According to Thompson (1988, p. 217) "Bayes Theorem shows that anything close to an accurate interpretation of the results of imperfect predictors is very elusive at the intuitive level. Indeed, empirical studies have shown that persons unfamiliar with conditional probability are quite poor at doing so (that is, interpreting ratings results) unless the situation is quite simple." It seems likely that the combination of less than perfect data with less than perfect users could quickly yield completely unacceptable practices, unless safeguards were in place to insure that users knew how to recognize problems of validity and reliability, understood the inherent limitations of rating data and knew valid procedures for using ratings data in the contexts of summative and formative evaluation. (79-80).

The authors conclude by noting "It is hard to ignore the mounting anecdotal evidence of abuse. Our findings, and the evidence that ratings use is on the increase, taken together, suggest that ratings malpractice, causing harm to individual careers and undermining institutional goals, deserves our attention." (pp. 79-80).

Then in endnote #26 of my paper, Theall objects to my quoting Dowell and Neal noting that: "...Cohen's (1983) review reinforces our earlier conclusion that student ratings are inaccurate indicators of student learning..." (Note 2) Theall omits the rest of the Cohen's quote which said, "and therefore are best regarded as indices of consumer satisfaction' rather than teaching effectiveness." Even given that Theall (et al.) is talking about the logic in use of SEF, not its inherent validity, he goes on to say,

> Conversations with faculty and administrators...led increasingly to concerns about what users [e.g., chairmen; deans] were doing with the information we were providing. We saw that some departmental administrators, who routinely use ratings to make decisions about personnel, evaluation policy, and resource allocation, were not familiar

> enough with important ratings issues to make well informed decisions...Clearly stated disclaimers regarding the limitations of ratings data in particular circumstances appeared to have little effect on the inclination of some clients to use invalid or inadequate data...There are some fundamental concepts for using numbers in decision making. To the degree that these concepts are ignored, interpretations of data become, at best, projective tests reflecting what the user (e.g., a chairperson or dean) already knows, believes, or perceives in the data. Treating tables of numbers like inkblots ('ratings by Rorschach') will cause decisions to be subjective and liable to error or even litigation...

Theall (et al.) continue in that chapter to lay out carefully the methodological precondition for the reliable use of SEF . For example, in that chapter Theall's (et al.) wrote:

> Three types of errors come to mind immediately. The first involves interpretation of severely flawed data, with no recognition of the limitations imposed by problems in data collection, sampling, or analysis. This error can be compared to a Type I error in research -- wrongly rejecting the null hypothesis -- because it involves incorrectly interpreting the data and coming to an unwarranted conclusion. In this case, misinterpretation of statistics could lead to a decision favoring one instructor over another, when in fact the two instructors are not significantly different...(p.87-88)
>
> The second type of error occurs when, given adequate data, there is a failure to distinguish significant differences from insignificant differences. This error can be compared to a Type II error. -- failure to reject the null hypothesis - because the user does not realize that there is enough evidence to warrant a decision. In this case, failure to use data from available reports (assuming the reports to be complete, valid, reliable, and appropriate) may be prejudicial to an instructor whose performance has been outstanding but who, as a result of the error, is not appropriately rewarded or worse, is penalized. (p.87-88)
>
> The third type of error occurs when, given significant differences, there is a failure to account for or correctly identify the sources of differences. This error combines the other two types and is caused by misunderstanding of the influences of relevant and irrelevant variables. In this case, a personal predisposition toward teaching style.., may lead a user to attribute negative meanings to good ratings, or to misinterpret the results of an item as negative evidence when the item is actually irrelevant and there is no quantitative justification for such a decision. Any of these errors can render an interpretation entirely invalid.(p.87-88)
>
> Summarizing, Theall (et al.) wrote,
>
> Let us...state our goal in the following way: "The user will make decisions that are based on valid, reliable hypotheses about the meaning of data." In this case, the user should receive or construct working hypotheses that do the following things:
>
> - Take into account problems in measurement, sampling, or data collection and include any appropriate warnings or disclaimers regarding the suitability of the data for interpretation and use.
> - Do not attempt to account for differences between any results when they are statistically not significant (probably <.05).
> - Disregard any significant differences that are merely artifacts (for example, small

> differences observed in huge samples), which can technically be significant but are unimportant).
>
> - Account for any practically important, significant differences between results in terms of known, likely sources of systematic bias in ratings or reliably observed correlations, as well as in terms of relevant praxio logical constructs about teaching or instruction.
> - The user should also refrain from constructing or acting on hypotheses that do not meet these conditions...(pp. 87-89)...

Again, I said in my paper,

> SEF may indeed validly describe incompetence...given (a) the conflicting data on their validity, (b) the way many institutions have constructed SEF instruments, (c) the often unsystematic statistical method by which SEF are interpreted, and especially (d) given the considerable weight accorded negative comments by only a few students in making tenure and promotion decisions, it would seem SEF can all to easily be used as a covert instrument for the elimination of tenure candidates and other faculty who may threaten student tuition dollars and perhaps ideological and popular culture agendas.

So what is the problem of "intent" to which Theall refers?

Inappropriate Tone of Theall's Commentary

Theall's also makes brief mention of Stake's (1997) response to my article. While Stake can take care of himself, I would like to comment on one of Theall's comments as in my view it is indexical of Theall's setting up "straw men." Theall says,

> Stake's reference is probably to "educational seduction", the skill of the infamous "Dr. Fox" who supposedly entertained students and received high ratings despite the fact that he delivered no content (Naftulin, Ware, & Donnelly, 1973). Many who do not care for ratings find one study that supports their position but ignore subsequent work (e.g. Perry et al., 1979; Marsh & Ware, 1982) which points out problems with the original study and proceeds to clarify the issue.

The point is that Stake does not even come close to mentioning the Dr. Fox study, a study that used an actor to deliver lectures with nearly no content. Setting up such straw men, allows Theall to then claim to have demolished them. What the purpose of this was, I can not say.

Finally, in a critique which rather cavalierly takes on not only my article, but Stake (1997) and Greenwald, and Greenwald and Gilmore (1996), (Note 3) perhaps what is most disturbing is that Theall's attempt at rebuttal is, in tone, unjustifiably condescending and, in word, *ad hominem*. In referring to a research article of his (et al.), from which I quote in my article much to his dissatisfaction, he says,

> Frankly, I don't like to recommend articles like this to those not actively involved in ratings research or practice because such writings can mislead readers who aren't really familiar with the cited ratings literature. I'm sorry if that sounds elitist: it isn't intended to be but I do have a reason for noting it.

Theall should be more than "sorry;" as a scholar presumably knowledgeable in science, he should be embarrassed. He should also not flatter himself by suggesting his statement may be considered elitist; arrogant, perhaps, but not elitist. (Note 4) This is why: First, he has no legitimate reason to assume that I am not familiar with the literature (see below), just because I disagree with him and have reservations about SEF . His assumption seems to be based on the most fundamental of logical and statistical inference fallacies. Theall says,

> We (Franklin & Theall, 1989) found that ignorance of evaluation/measurement literature and methods correlated significantly with negative faculty opinions about students and student ratings. I note this because discussions about ratings are so often filled with misinformation.

Now as any first year college student learns in statistics, (a) correlation is not causation, and (b) one can not infer from a statistical generalization of a group, that an individual in that group is an indicator of that generalization. In fact this "logic" is the same as that used by people who exhibit prejudice: applying some generalization or central tendency characteristic of a specific group of people to an individual. Scriven (1988) makes this point in a similar context.

Second, were his claims not so clearly uninformed, they would be an insult not only to me, but to many other scholars as well. This is why that is: To say that scholars not actively involved in doing statistical or experimental research on a subject are not qualified to read, understand and comment on this research, immediately puts most theoretical physicists, theoretical biologist, and philosophers of science out of business. Hmmm. (Note 5)

References

Barnett, Larry, D. (1996). Are Teaching Evaluation Questionnaires Valid? Assessing the Evidence. Journal of Collective Negotiations in the Public Sector, 25, (4) 335-349

Cohen, P. A. (1983). Comment on a selective review of the validity of student ratings of teaching. Journal of Higher Education, 54, 448-458.

Franklin, J. L. & Theall , M. (1990). Communicating student ratings to decision makers: design for good practice. In M. Theall & J. Franklin (Eds .) Student ratings of instruction: issues for improving practice. New Directions for Teaching and Learning # 43. San Francisco: Jossey Bass.

Franklin, J. L. & Theall, M. (1989). Who reads ratings. Knowledge, attitudes, and practices of users of student ratings of instruction. Paper presented at the 70th annual meeting of the American Educational Research Association. San Francisco: March 31.

Greenwald, A. G. & Gillmore , G. M. (1997). No pain, no gain? The importance of measuring course workload in student ratings of instruction. Journal of Educational Psychology (in press).

Greenwald, A. G. & Gillmore , G. M. (1996). Applying social psychology to reveal a major (but correctable) flaw in student evaluations of teaching. University of Washington, Draft Manuscript, March 1.

Haskell, R. E. (1997b). Contributed Commentary on: Stake Response to Haskell: Academic Freedom, Tenure, and Student Evaluation of Faculty. Educational Policy Analysis Archives, 5 Available online: http://olam.ed.asu.edu/epaa/v5n8c1.html.

Haskell, R. E. (1997a). Academic freedom, tenure, and student evaluations of faculty: Galloping polls in the 21st century. Educational Policy Analysis Archives, 5. Available online: http://olam.ed.asu.edu/epaa/v5n6.html.

Natuflin , D. H. Ware, J. E., & Donnelly , F. A. (1973). The Dr. Fox lecture. A paradigm of educational seduction. Journal of Medical Education, 48, 630-635.

Scriven , M. (1988). The New Crisis in Teacher Evaluation: The Improper Use of 'Research-based' Indicators. Professional Personnel Evaluation News. (January) 1-8.

Stake, J. E. (1997). Response to Haskell: academic freedom, tenure, and student evaluations of faculty. Educational Policy Analysis Archives, 5. Available online: http://olam.ed.asu.edu/epaa/v5n8.html.

Theall , M. (1997). On Drawing Reasonable Conclusions about Student Ratings of Instruction: a Reply to Haskell and to Stake. Educational Policy Analysis Archives, 5. Available online: http://olam.ed.asu.edu/epaa/v5n8c2.html.

Thompson, G.E. (1988). Difficulties in interpreting course evaluations: Some Bayesian insights. Research in Higher Education, 28, 217-222.

Notes

1. Theall, suggested that I misinterpreted research on the SEF. For example, he says,

> Peter Cohen's most recognized contribution to the ratings literature is the meta-analysis of multisection validity studies in which performance on a common final exam was correlated with ratings (1981). There was a .43 correlation between ratings and performance on those exams. In other words, sections with highly rated instructors had higher average scores which could not be attributed to differential grading standards or sampling errors. This is evidence for the validity of ratings, the exact opposite of what Haskell says. Calling that relationship grade inflation is simply ridiculous!

Further, in at least partial defense before readers who may not be conversant with the literature, let me offer at least one somewhat lengthy statistical anecdote from Barnett (1996) that addresses Theall's belief in .43 correlations. In analyzing some of Cohen's (1981) data, Barnett explains,

> I turn now to a second statistical matter the interpretation of correlation coefficients. The numerical value of a correlation coefficient can be deceiving, even when it is 'statistically significant' (i.e., even when it is unlikely to have occurred by chance if no relationship exists). A correlation coefficient measures the degree to which change in the amount of the explanatory variable is accompanied by change in the amount of the effect variable, but the most beneficial feature of a coefficient is not its numerical value, which has no inherent, practical meaning. Rather, the square of the numerical value is the most advantageous aspect of a correlation coefficient, for the square indicates the proportion of variation in the effect variable that can be statistically attributed to variation in the explanatory variable. The research summary by Professor Cashin reported a correlation coefficient of .44 between student ratings of an instructor 'overall' and examination grades [8]. This coefficient means that 19.4 percent of the variation in student learning (as measured by course grades) is explained by variation in instructional quality (as measured by student ratings). If accurate, a correlation of this magnitude is 'practically useful,' as Professor Cashin said, though one must keep in mind that four-fifths of the variation in course grades remains unexplained and is attributable to other factors.

But does this correlation coefficient accurately estimate the relationship between student evaluations of teaching and student achievement? The best research on the magnitude of the relationship is the ' multisection validity study.' When it is ideally designed, such a study possesses the following features: each course included in the study has numerous sections; students are randomly assigned to sections; the sections of a course have different instructors but a common textbook and the same examination(s); all examinations for a course are constructed by a person who does not teach any section; and subjective (essay) components of examinations are graded by the person who developed them. A review of multisection validity studies cites one work that, the author of the review asserts, eliminates at least in part 'many of the criticisms of the multisection validity study' and ' provide[s] strong support for the validity of students' evaluations of teaching effectiveness' [9, p. 721]. However, the cited work which subjected the results of other multisection studies to a statistical analysis did not control a number of critical variables that could have generated or enlarged the relationship between student ratings of teachers and student achievement [10]. Among the missing variables that might have explained the relationship was the rigor of the requirements of the instructor (such as checks for student preparation and amount of material assigned), a factor that may vary considerably across sections of a single course. If the variable was related both to student ratings of instructional quality and to student achievement, a control for the variable could have markedly weakened or entirely eliminated the relationship originally found between student ratings and student achievement p.339

Another variable that the work omitted was the students' level of interest in the subject matter of the course prior to exposure to the teacher they later evaluated. As will be suggested below, neither of these variables should have been excluded from the analysis and left uncontrolled.

While the work did not incorporate a number of potentially important variables into its data analysis, the work is the source of a set of correlation coefficients (including the coefficient of .44) that Cashin suggested are credible estimates of the relationship between student ratings of teachers and student achievement. A reader of the reproduced coefficients can easily be misled, however, because Cashin failed to make clear that the coefficients may have been seriously confounded by variables whose influence was not removed. The failure to clarify this point is surprising inasmuch as Cashin explicitly stated that a control is necessary for one of the variables omitted by the work, namely, the interest students initially exhibited in the subject.' [12, p. 5]. P.340

Cohen treated student judgments of 'the amount and difficulty of the work the teacher expects of students' as one component of instructional quality. Whether it is an element of teaching quality or a separate factor, the amount and difficulty of work should be controlled under the conditions mentioned because it may explain much or all of the relationship detected between, on the one hand, evaluations of an instructor overall or on specific dimensions and, on the other, performance on examinations. From the studies he reviewed, Cohen calculated a negligible mean correlation coefficient for the relationship between the amount/difficulty of work and student achievement, but he also found a substantial range for the coefficients reported by the studies. Specifically, the interval for 95% of the coefficients extended from -.42 to +.39. Id, at 293, 295. Individual studies may thus involve a nontrivial association between the perceived difficulty of teachers and the examination performance of students (p.341).

I would also suggest that he reader see Greenwald's historical review of past studies and

Barnett's review of some of the most cited pieces that claim the validity of SEF.

2. I can understand how such quotes might be embarrassing (a) to an author who professes that SEF is basically if not completely valid instrument, and (b) to a practitioner trying to convince faculty and administrators to use them in faculty evaluation programs. I have never heard of a consultant who suggests that SEF have been shown to be invalid. Given Theall's critique let me make it very clear that I am not here indicting consultants. I have done and occasionally do consulting myself. I would like to say that Theall (et al.)---in word, I can not speak for deed---does recognize most of what needs to be done to make SEF use in administrative decisions reasonably ethical.

3. Requests for prepublication copies of Greenwald's articles should be sent to http://weber.u.washington.edu/~agg/

4. Theall goes on to say, "Stake says that ratings undermine the faith and trust students must place in teachers." Given his opinion of ratings, I assume that he must then feel that teachers have no need to place faith or trust in their students. This paternalism is antiquated and unrealistic. The logic by which it is unacceptable for Stake not to trust (valid?) perceptions of students, but acceptable for Theall not to trust the perceptions of scholars, escapes me, just a as does his calling Stake's view of students paternalistic and not seeing that his own views of his not trusting scholars to correctly read research is paternalistic. Though he does recognize his view sounds elitist, he says it is really not.

5. I wonder what Theall would say about child psychologists who have no children of their own, or about healthy psychotherapists who treat the mentally disordered.

EPAA
• enter the archives • browse the abstracts • the editors • the edit board
• submit article • submit commentary • search • subscribe
volume: • 1 • 2 • 3 • 4 • 5 • 6 • 7

This article has been retrieved 3087 times since February 25, 1997.

Education Policy Analysis Archives

Volume 5 Number 7 **February 25, 1997** **ISSN 1068-2341**

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass Glass@ASU.EDU.

College of Education

Arizona State University,Tempe AZ 85287-2411

Copyright 1997, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

The Circle of Learning: Individual and Group Processes

Ernest Chang
Axia Multimedia Corporation

echang@axia.com

Don Simpson
Axia Multimedia Corporation

Abstract:

We present a paradigm for modeling the processes found in individual and group learning. Using combinations of two dimensions, the first being whether the learner's *activities* are By-Oneself or With-Peers, and the second whether the process *orientation* is toward the Person as the focus of the learning or toward the Group as the focus, we derive four quadrants in Activity-Orientation learning space.

These four quadrants represent: lectures, individual learning, concurrent learning, and collaborative learning. From these combinations of Activities versus Orientation, we can describe many characteristics of these different learning *categories*.

Table of Contents

Introduction
The Activity-Orientation Paradigm
Four Distinct Learning Categories
    Quadrant A: Traditional Lectures
    Quadrant B: Self-Study
    Quadrant C: Concurrent Learning
    Quadrant D: Collaborative Learning
Multi-Dimensional Attributes of the Model
    The Interpersonal Dimension
    The Learning Environment
    Knowledge Content
    Technology Support for Learning
    Sociological Dimensions
The Circle of Learning
Discussion
Conclusion
References

Introduction

Traditionally, many educators have considered learning to be an individual responsibility, with students accepting the burden of acquiring knowledge and expertise . Recently, the notion of collaborative learning has been strengthened, from a number of sources. These include the push in Kindergarten to Grade 12 and colleges to learn through group projects within a classroom (Felder 1995; Johnson & Johnson 1991), and through communicating with other students across a network, such as KidNet from the National Geographic Society, or interactive video (Rettinger 1995; McArthy 1995) in the domain of distance learning. Digital communications networks such as the Internet Internet (Vetter 1995; Macedonia 1994), or the use of Lotus Notes, have become the new medium in which group learning is anticipated to take place, and many large businesses have already built internal group learning systems using Lotus Notes.

Organizations and businesses have increasingly moved to an understanding that in a continuously changing environment, the "learning organization" is the high-performance organization (Gordon 1992; Senge 1992). In a knowledge economy, the organization which adapts best through new knowledge by effective learning (Mumford 1993; Vowles 1993) is the one which will take the lead, and maintain it.

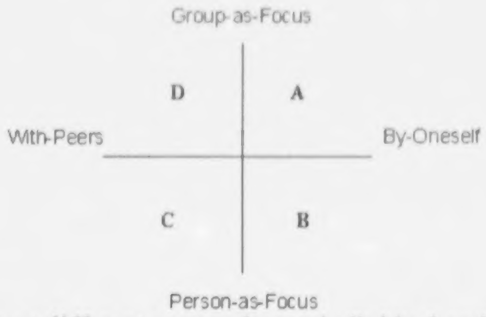
It is the purpose of this article to explore a simple paradigm for individual and collective learning that is inclusive of the many shades of meaning in this domain, and will serve to clarify the relationships between the several inter-related concepts. We will then explore the significance of this paradigm in terms of practical implications for future action.

The Activity-Orientation Paradigm

This is based on the observation that there are two dimensions along which learning takes place:

By-oneself vs With-Peers	Learning Activity
Person-as-Focus vs Group-as-Focus	Process Orientation

When taken together, they form a coordinate system with four quadrants:



These quadrants represent different approaches to individual and group learning across a number of dimensions. We will now explore these characterizations.

The Learning Activity Dimension

By-Oneself means that the learning process is one in which the student acts alone, even if physically he or she is with others. Studying by reading in one's own room or office is the protoypical example of learning By-Oneself. However, in an extended learning-by-being-told situation, such as a lecture, or briefing, the student is, in terms of Activity, still By-Oneself.

On the other end of this dimension is learning With-Peers, in which learning activities involve extensive or continuous interaction with others. The typical university seminar, most junior grades in high schools, the study group, the conference call, or the computer-supported conference are examples of learning With-Peers.

The Process Orientation Dimension

We consider learning as a process within a social context, in which more than one person may be present (by process, we mean a sequence of activities directed towards a specific goal). In this context, the process can be oriented either towards the Person-as-Focus, or the Group-as-Focus. As this process, by definition, involves all participants, the orientation is independent of the perspective of the persons involved. For example, whether from the viewpoint of the student or the teacher, the process orientation in a lecture, in which the teacher stands in front of a class, is that of Group-as-Focus. In other words, the teacher's point of view is that the group is being addressed. The student's point of view, in terms of process orientation, is that the teacher is not addressing the student, but the entire group.

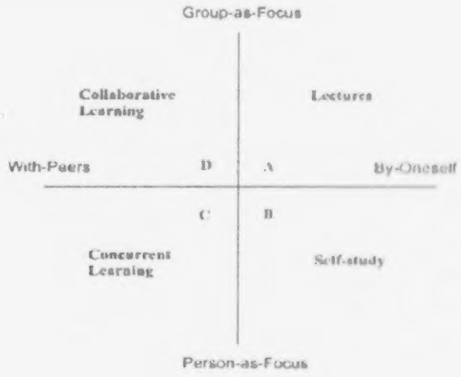
A learning process oriented to the Person-as-Focus is the student studying alone, of course. Another is that of an informal study group, or a typical internet newsgroup, in which the individuals interact with others, but the process is based on each person producing their specific contributions and meeting their individual needs.

On the other hand, if a learning process involves the group, in which some external goal or structure has been imposed, such as requiring the group to make a decision, or to come to agreement on a position regarding an issue, the process would be Group-as-Focus.

Table of Contents

Four Distinct Learning Categories

The distinct learning processes represented by the four quadrants correspond to distinct categories of learning, as shown below. In much of the literature on learning, they have been discussed in isolation, or in juxtaposition (such as group versus individual learning (Brown 1991; Carley 1992; Taylor 1992) or lectures versus collaborative learning (Ransbury 1994; Garko 1994; Johnson 1991) without an analysis of their structural relationships. We will first describe the general characteristics of each learning category, and then discuss the ways in which a number of important attributes differ for each.



Quadrant A: Traditional Lectures

In a traditional lecture in which a person talks while others listen, the student's activity is By-Oneself, while the orientation of the process is towards the group as a whole, ie, the Group-as-Focus. Although the wish is frequently made that students interact more in this process, studies of student questioning (Dillon 1988; Graesser 1994) reported that in the typical lecture situation, the frequency of questions per hour is only 3.0. The process dynamics of this learning category is that of a single person addressing the group as an entity, so it is no surprise that this process does not support With-Peers behavior well.

Where this learning category demonstrates efficiencies is that a single person can place the same information within reach of many. This is the time-honored way of utilizing the scarce resources of experts, reaching back to a time when there was only an oral tradition of passing knowledge from one person to others.

The lecture also has other potential strengths, including the ability to motivate others to certain behaviors, to inspire them to reflect deeply, or to perform logical analysis. By example, a lecturer can demonstrate the synthesis of complex information, or the integration of several ideas into a coherent whole.

Quadrant B: Self-Study

The learning activities in this learning category are those performed by a person By-Oneself (regardless of whether the person is physically alone or not). The process is also oriented to the person as the focus. Thus, solo reading, problem solving, individual experiments are typical of this category.

However, some learning processes in this category could involve more than one person. For example, when a tutor working with a student, the learning is being done by that student, and all the processes are also focused on the student.

It follows, from the above, that the technology support for Self-study includes both electronic media as book (for the activity of solo reading), and as electronic tutor.

Quadrant C: Concurrent Learning

The learning activities in this category are in the character of being With-Peers, in the sense that the activities involve learning together, through shared activities in a collegial manner. The processes are oriented around the Person-as-Focus, in that the goals and outcomes of the social and learning processes are individual in nature, rather than focused on the group as a whole.

Typical of this model of collaborative learning are group interactions such as study sessions, Internet newsgroups, and ad hoc project groups, in which peers express and exchange opinions, values, perspectives and some facts, or accomplish common objectives that satisfy individual goals.

Quadrant D: Collaborative Learning

In contrast to Concurrent Learning, this learning category concerns those processes oriented to groups as entities. The perspective is that of the goals and outcomes of the group, using measures which reflect the group as the entity under consideration. The individual's activities are in the With-Peer framework, in that he or she participates in processes which are highly interactive and collaborative in nature.

The differences between Concurrent Learning and Collaborative Learning are the differences between the collection of goals of individuals, and the goals of a coherent collective culture, which can be viewed as an entity with its own measurable goals, achievements and outcomes. The behavior of both are based on With-Peers activities, in other words, on collegial interactions. With Concurrent Learning, a group of peers interact to achieve their individual learning goals; with Collaborative Learning, a group of peers interact to achieve their collective learning goals.

This distinction may appear to be based only on the difference between having a stated objective for peer interaction or not, but it is more than that. For a group of persons to be oriented to the Group-as-Focus requires an adoption of common goals, values and culture that are coherent and persist over time.

Table of Contents

Multi-Dimensional Attributes of the Model

The four learning categories: Lectures, Self-Study, Concurrent Learning and Collaborative Learning can be further described by a number of common attributes, which serve to additionally characterize them. From these attributes we can draw a number of observations of a practical nature, concerning their inter-relationships and their implementations.

The attributes that will be used are:

- the interpersonal dimension

- the learning environment
- the knowledge content
- technology support
- sociological dimensions

The Interpersonal Dimension

The four learning categories described above differ significantly in the way in which the learner is involved with oneself and others.

In Lectures, the individual is listening to the lecturer, in a passive rather than interactive way; the listener is the recipient, the lecturer is the active dispenser of knowledge. The listener is apart from others who may also be present, and in an interpersonal sense, they may as well not be present. The lecturer often is the source of authority that validates the process.

Self-Study is usually an individual activity, focused on internal cognitive processes. These learning activities often take place in the isolation of one's own room, office or laboratory. Where others are present, such as in a library, special social and interpersonal rules are established which provide virtual isolation. The primary motivator for the Self-Study process is usually the student.

A special situation of Self-Study in which another person is involved in a meaningful way, is that of the Tutor who works one-on-one with a student. The learning activities are all directed at the student's needs, and fit into the Person-as-Focus definition; there are no peers interacting with the student, and therefore the process orientation is also By-Oneself. The interpersonal relationship with the tutor is an interactive one, where the tutor is in a sense, a dynamic extension of the knowledge to be acquired, providing guidelines, highlights, key strategies and motivation.

Concurrent Learning describes situations in which a learner is interacting with peers, but the learning activities are individualized, rather than having a group focus. The nature of the interactions with others is that of exchanges of facts, opinions and values. There may often be a set of individual goals, rather than a common learning goal. Among common examples of Concurrent Learning are the seminar, classroom discussion, hallway or coffee chat at work. Included in this category are distributed interactions such as electronic forums, Internet newsgroups and bulletin boards. Thus, the style of interpersonal relationships may be competitive, confrontational or collegial, rather than consensus-reaching.

It is not unusual for Concurrent Learning situations to have either formal or ad hoc referees, who facilitate the interactions and maintain them at a reasonable level of focus, interest and civility. Electronic forums and bulletin boards use formal facilitators, while informal discussion groups find individuals taking on the role of the facilitator from time to time. The legitimacy of Concurrent Learning sessions are imbued as much in the general sense of acceptance of the process by the individuals, as in the existence and activities of the facilitators.

The Collaborative Learning category describes situations in which a group of persons share common learning goals, and work together to achieve them. In a broad sense, achieving any shared goal is a learning experience. For example, agreeing on a new design for an engine part requires the contributions and assent of different members of the group, each representing different aspects of manufacturing, materials and operations. The end result is that each person comes away with new knowledge of the topic, interaction process and of the participants. There is of course a narrower sense in which a group of persons cooperate to achieve specific learning goals, such as a group of high-school students studying the composition of soil, and writing a joint report.

From our point of view, the interpersonal dimensions are the same. The flow of activity and thought is not directed by any individual; each person offers contributions to the group, and interacts with the group as a whole rather than with other individuals. The difference between

Concurrent Learning and Collaborative Learning, in the interpersonal dimension, are characterized by differences in group process. Concurrent Learning is a forum of individuals; Collaborative Learning is an environment in which each person attempts to be a coherent part of a whole, synthesizing with one another a shared understanding of values as well as facts. The personal behaviors required of each person in such a setting are not necessarily intuitive or natural, and it is not uncommon for Collaborative Learning to be mediated and orchestrated. For example, Robert's Rules of Order is a very formal set of rules that represents one highly structured way of achieving group consensus, among a whole spectrum of other approaches that are available to help a group act and learn collaboratively.

The important point is that Collaborative Learning, like the other learning categories, is characterized by specific interpersonal behaviors which arise from differences in Learning Activity (by-oneself or with peers) and Process Orientation (individual as focus or group as focus).

To summarize, the characteristics of the four learning categories in terms of the interpersonal dimension are:

Lectures	listening
Self-Study	focused on own thoughts
Concurrent Learning	participatory
Collaborative Learning	cooperative

Table of Contents

The Learning Environment

These categories also differ in the environments they provide within which learning takes place. In a Lecture, a person is subjected to a continuous stream of information, with little time to reflect on any specific part of it, at the risk of losing what follows. The Lecture environment constrains the listener to conform to the tempo of the lecturer's delivery, and to diverge from that sequence of ideas minimally.

Learning in the Self-Study category is quite different. The student is in complete control of what is done next, mentally and physically. There is freedom to contemplate the relationship between two concepts, to explore a thought association, to work out something that was not being well understood. This self-directed environment requires discipline and focus, and for some persons who do not have this rigor, is an inefficient way of spending time and effort in learning.

The Concurrent Learning category is one in which the learning process is individually focused, while the activity is with peers. This environment is one in which many participants express their individual opinions, beliefs, and arguments, in an open forum in which the competing ideas of the persons involved shift into and out of focus. The stimulation of receiving many different perspectives from others is offset by the inability to think much about any one thing before either getting another, or taking some action oneself. The environment is characterized by debate, and the validation or refutation of the arguments and supporting facts of the participants is very different from Lectures and Self-Study.

In the Collaborative Learning paradigm, both learning focus and activity are oriented to the group. The individuals in the group work together to achieve common learning goals, arriving at consensus. The environment is characterized ideally by sharing, openness, acceptance of the contribution of others, and the development of a cohesiveness in which each person becomes aware of the shared achievement of all. In this sense, successful Collaborative Learning among team members in a company is "organizational learning", with the organization having learned if its members have acquired shared knowledge and values in both a self-aware and a group-aware way.

In summary, the learning categories provide different learning environments:

Lectures	constrained to cognitive tempo of the lecturer
Self-Study	self-directed internal and external processes
Concurrent Learning	open forum for competing priorities and values
Collaborative Learning	consensus seeking based on common goals

Table of Contents

Knowledge Content

What is being learned in each learning category? Although one could argue that anything can be learned in any of these processes, and this in fact occurs all the time, we believe that their characteristics lend themselves to different kinds of knowledge. There is therefore a sense in which the learning category has a primary knowledge function.

As we have observed, there is a stream of information that flows from the speaker in the Lecture situation, so that the learner is constrained to follow the sequence and tempo of the lecturer. This reduces the opportunity for extended reflection and integration of new knowledge, while facilitating the introduction of facts, concepts, relationships, values, etc. in a more superficial way. The Lecture is better-suited for guiding the learner to appreciate the framework of a subject, its key concepts and highlights, rather than to a detailed understanding.

This is not to say that a lecturer is unable to lead the student, through a clear sequence of steps, to a more profound and deeper understanding of a specific concept than the learner could achieve by him/herself. Rather, the constraints of the leader-follower relationship lend themselves to the lecturer imparting knowledge to the listener. When there is a group in attendance, the lecturer will tend to address the common level of understanding of the audience, which in general reduces further the complexity of the delivery.

In Self-Study, the learner follows his or her own personal initiative in working with the material. Whether reading a book, watching a video, doing exercises, writing, listening to a tape, the learner is in full control of what next to do or think about. If a concept is not well understood, or sparks a link to a new idea, the learner can choose to pursue the issue further, to whatever extent is desired. This personally directed, potentially non-linear flow of internal events means that the learner can seek to achieve a mastery of the subject matter, in a way that is not possible in the Lecture or the other interactive group-focused categories. The kind of knowledge that is realizable is therefore not only factual, but also the associations, relationships, and use of these facts in linkages to the learner's existing knowledge contexts. For this reason, Self-Study, which includes reflection, assimilation, integration and association of new concepts, as directed by the learner, is essential to furthering a person's learning in most domains.

Concurrent Learning describes an environment in which individuals pursue their own learning goals while interacting with one another. The pace and tempo of this interaction is not determined by any one person, and therefore reflection and self-directed thinking is not the order of the day. Rather, the learner has the opportunity to weigh a number of positions, opinions and arguments proferred by others, and in turn to construct and propose his or her own contributions. This open forum of competing rhetoric provides the learner with a unique kind of learning, in which one's own knowledge is seen in the perspective of others, and in which one can appreciate the same topic from several points of view. In order to participate, the learner has to generate coherent knowledge structures dynamically, which reflect linear traces (because voice and written language is linear) through the internal network of concepts and relationships. The dynamic coherence of this output, relative to the peer-based demands of the current conversational context, is a reflection of the facility with which a person has conscious mastery of specific knowledge. In fact, the dynamic creation of such coherent structures is a generator of expertise.

In Collaborative Learning, individuals work together to achieve common learning goals, which are often declared formally, but may also be implicitly assumed in the process. The consensus that leads to group decisions are based as much on an understanding of shared values as on the set of facts and rational constructs through which these values find expression. This shared awareness is in a very real sense group learning, and is one of the keystones of organizational learning.

To summarize, the learning categories differ in the types of knowledge that are acquired:

Lectures	speaker imparts knowledge to the recipient
Self-study	self-directed reflective integration of subject
Concurrent Learning	generate own knowledge in perspective of others'
Collaborative Learning	group consensus based on shared values

Table of Contents

Technology Support for Learning

There is general concern that traditional learning approaches may be inadequate in the face of many pressures: the increasing quantity and quality of learning demands, the increasing diversity of the student group, and the decreasing amount of time available for learning needed to address changing situations. What technology support offers to each learning category is the opportunity to individualize or make more effective the mechanisms through which learning takes place.

Lectures are limited in physical size by room space, and in audio-visual space by the audibility and legibility of the lecturer's voice and visuals. These can and have been enhanced in many ways: amplification, screen projectors, live video into several rooms, live video over a broadcast, satellite or private network, and recorded video and notes for distance learning in one of its modes.

Self-study is a learning process in which students work at their own pace to acquire, reflect upon and integrate knowledge. Technology support for this learning process enhances this in a number of ways. It can provide the student with more powerful representations of knowledge, such as sounds, videos, and images. More important, learners have the freedom to move in this multimedia knowledge space that is almost as powerful as the freedom of self-directed thought. Facts, ideas, opinions are no more than a few keystrokes or mouse-clicks away. Most of all, technology support can provide an interactive environment in which the learner can find support, guidance, and responsiveness that is close to what a human tutor offers. In this way, a learner is able to focus on personal learning goals through a supportive, adaptive and guiding process, that is at the same time highly personalized. This kind of technological support for learning in the Self-Study process is highly empowering, promising large potential gains in learning effectiveness and efficiency.

The technology support for Concurrent Learning has been directed at increasing the scope and power of the open forums in which these learning processes take place. To increase the numbers of persons involved, video conferencing or computer conferencing can be used. The storage and processing power of the computer makes it possible for conferences to be either synchronous (at the same time) or asynchronous (whenever one gets to it). The messages between persons can either be documents, text, voice or full video. Many of these conferences are moderated by a coordinator, but this does not necessarily improve their usefulness, depending on one's point of view. Many distance education programs use technology support of this kind as the backbone through which information is disseminated, and discussions are held. As an example, the Center for Innovation and Management of Athabasca University uses this technology, through Lotus Notes, for all of its MBA program, delivered and implemented through distance learning.

Collaborative Learning is an environment in which a group of persons participate in a learning process that has common goals. The technology support that has been developed in support of these processes has been called Group Decision Support Systems (GDSS). Typically, a computer network supports a group of persons who sit in a face-to-face environment, each person their own computer display and keyboard also in front of them. Through the system, individuals move through a structured but flexible series of interactions that serve to highlight the issues involved, the values that are expressed, and provide processes for resolving differences and coming to consensus (Watson 1988, Nunamaker 1991).

The four categories of the Activity-Orientation model of learning differ in their characteristics. Not surprisingly, the need for, and the nature of the technologies that support these learning categories vary:

Lectures	simulcasting; recorded videos
Self-Study	interactive multimedia
Concurrent Learning	computer conferencing
Collaborative Learning	group decision support systems

Table of Contents

Sociological Dimensions

By this attribute, we refer to the relationships between the learner and the elements of the learning process in two ways: first, the basis for the interactions in terms of social validation, and second, the nature of the learning categories in terms of group dynamics.

For the learning process to be effective, the learner needs to believe that the process is a valid one. This validation is based on the student's relationship to the process, and these relationships are:

Lectures	authority
Self-Study	self-respect
Concurrent Learning	mutual respect
Collaborative Learning	shared values

In terms of the dynamics of the group (or individual) process, reflect on how the individual is empowered relative to the process. The categories can be characterized as follows:

Lectures	autocracy
Self-Study	autonomy
Concurrent Learning	democracy
Collaborative Learning	community

The structural relationship of the student to the learning category is therefore along two dimensions. The first is based on social validation, which leads to the second, based on empowerment.

The Lecture, for example, is a learning process in which the student's basis for validation is the authority of the Lecturer; this leads to the group dynamics of an autocracy.

For Self-Study to be effective, the student's must believe that it is a valid process for learning. This validation is based on self-respect. To the extent that the student has a strong sense of self-respect, the effort spent in Self-Study is associated with positive expectations; this leads to highly autonomous behaviors on the part of the student (as opposed to highly dependent behaviors for learning, in which the person only works if directed by the teacher or the group).

Concurrent Learning, at its most positive, facilitates individual learning through an appreciation of the perspectives of others, which lead to growth in one's own values, methods, goals and even concepts. For this to happen, mutual respect must be present; the group dynamics that reflect mutual respect is that found in democracies.

On the other hand, Collaborative Learning is successful when group-decision making is adaptive and leads to increased cultural coherence. This depends on individuals having, not just mutual respect, but a common membership in the collective culture, which we represent in the term "shared values". The type of group dynamics for people who have shared values are those of communities, in the sense that reside in the word "communitas"....a group of people who have shared values and behave coherently, as a group.

Table of Contents

The Circle of Learning

These attributes of the Activity-Orientation model of learning can be summarized in the following diagram, in which the characteristics, as listed below, are shown as layers of the model:

Sociological Dimensions
Technology Support
Knowledge Content
Learning Environment
Interpersonal Dimensions

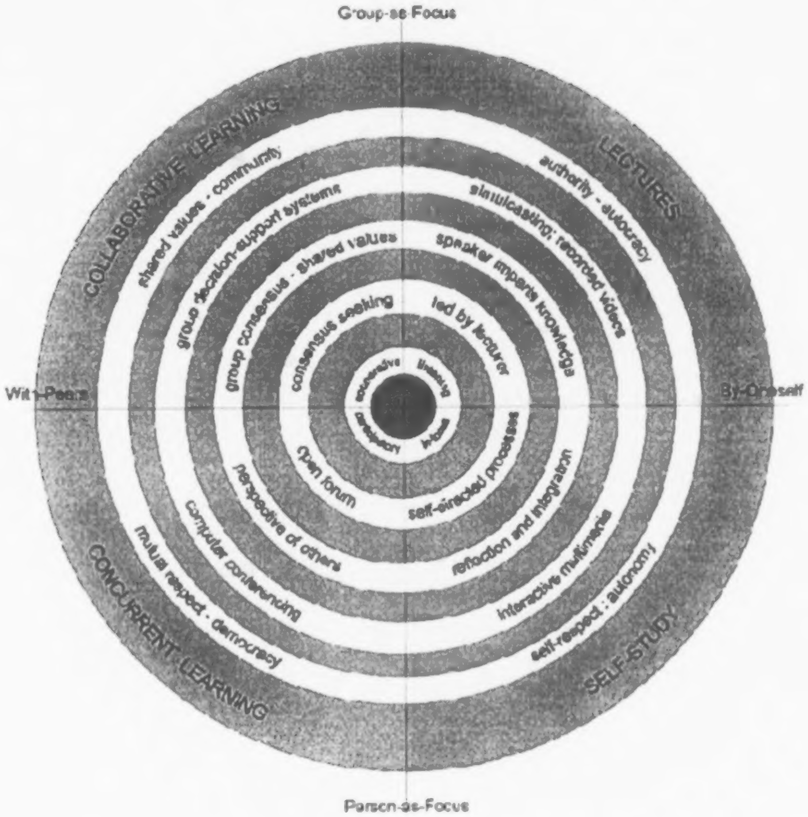

Group-as-Focus
COLLABORATIVE LEARNING
LECTURES
shared values - community
authority - autocracy
group decision-support systems
simulcasting; recorded videos
group consensus - shared values
speaker imparts knowledge
consensus seeking
led by lecturer
listening
With-Peers
By-Oneself
open forum
self-directed processes
perspective of others
reflection and integration
computer conferencing
interactive multimedia
mutual respect - democracy
self-respect - autonomy
CONCURRENT LEARNING
SELF-STUDY
Person-as-Focus

Figure 1. The Activity-Orientation Paradigm

Discussion

The four learning categories developed above originate from a two-dimensional paradigm, which as always is a simplification of situations found in reality. The purpose of the paradigm is to isolate and bring into relief typical characteristics of the learning categories, so that learning processes can be more easily understood within the complex environment in which everyday activities exist.

However, within a specific learning event, bounded by a particular group of people and a particular interval of time, it is very likely that a mix of the processes described in the four learning categories will occur. These will vary in intensity as well as in their order of appearance, depending on the characteristics of the learning event. The following are examples of the application of the paradigm to learning scenarios. Figure 2 shows the processes that would be expected in a typical classroom today in North America, where most of the learning transactions are through the lecture process, and a certain amount of reflective thinking goes on for individuals who disengage from the lecture or group process. The class may include a significant amount of collaborative learning, in which discussion takes place among students, or students work in groups, either mediated by the teacher as facilitator or independently. Some of the collaborative activity may be in the form of exchange of perspectives between students or student and teacher, but the group or sub-groups may also be working toward shared goals, as in the creation of a single project report, or the development of a common strategy for a business.

BEST COPY AVAILABLE

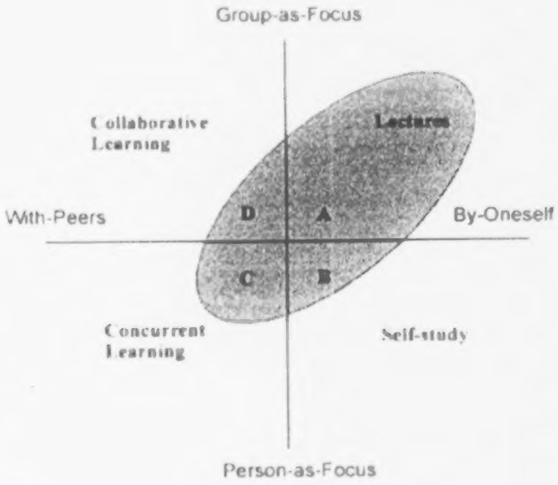
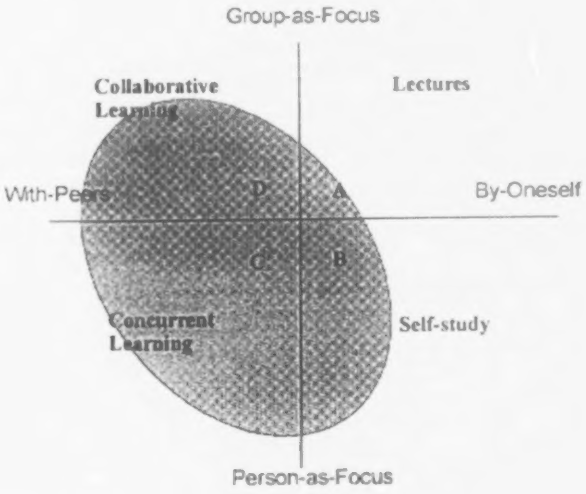

Group-as-Focus
Collaborative Learning
Lectures
With-Peers
D
A
By-Oneself
C
B
Concurrent Learning
Self-study
Person-as-Focus

Figure 2. Classroom Session

It is quite easy for a classroom in today's school to move fluidly between Lecture and group learning processes. In many schools, formal lectures are being replaced to varying extents by cooperative learning in small groups, which might be a situation as shown in Figure 3.

There may be an element of Lecture, in which the teacher provides the frame of reference for group activities. Most of the activities are in the With-Peers mode, divided between processes oriented to the self and group objectives. An example of the former is each student discussing his or her research on different parts of a problem; an example of the latter is coming to agreement on the conclusions of an experiment they conducted jointly. There is, within this environment, opportunity for a person to step back from group activities to work or think alone. However, the group usually exerts pressure to limit the amount of time-alone that has not been agreed upon.


Group-as-Focus
Collaborative Learning
Lectures
With-Peers
D
A
By-Oneself
C
B
Concurrent Learning
Self-study
Person-as-Focus

BEST COPY AVAILABLE

Figure 3. Cooperative Learning in Classroom

In Figure 4, we show the processes that would be found in a typical informal discussion session. Most of the overall effort is in the exchange of perspectives with an individual orientation. There may be some engagement in individual thinking during the session, and in the pursuit of overall group objectives or consensus.

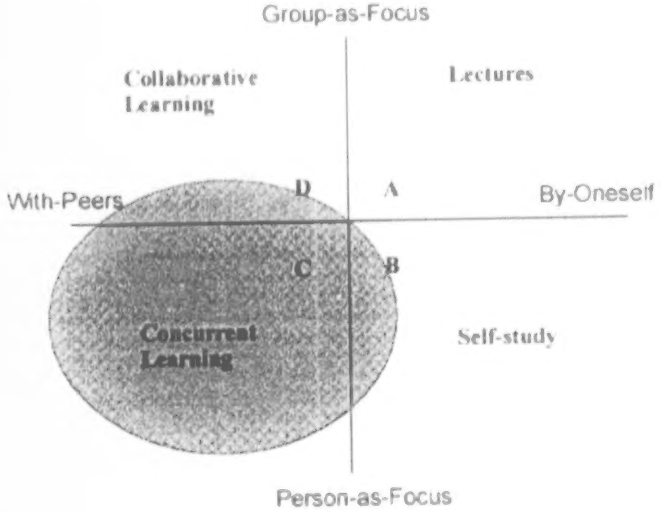


Figure 4. Informal Discussion Session

Figure 5 shows the learning processes that we would expect to encounter in a computer-mediated asynchronous conference over a period of time, such as occurs in the use of Lotus Notes to support courses. While most the effort goes into the exchange of perspectives and opinions (and possibly facts) between students, a good deal of reflective thinking and integration by individual students is made feasible by the fact that messages are read and replies originated off-line. Therefore, this session differs from informal discussion in that more Self-Study processes occur, in which knowledge integration is achieved through reflective thinking.

BEST COPY AVAILABLE

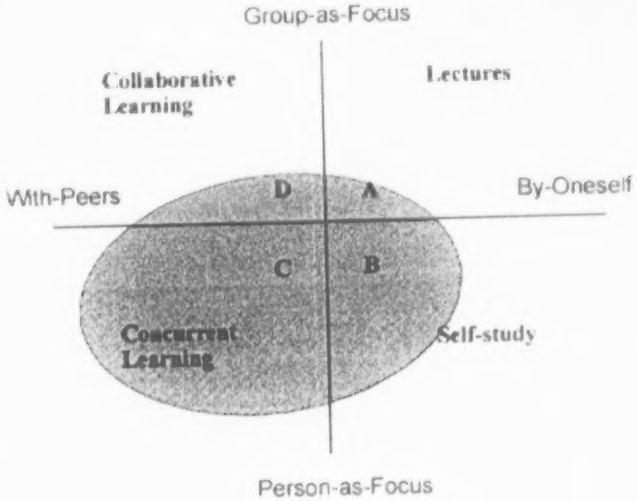


Figure 5. Asynchronous Conferencing

Figure 6 shows the mix of processes that might be found in a typical business meeting, in which peers interact to reach consensus on specific organizational goals. Some didactic information-giving typically starts the meeting off, and then a significant amount of exchange of perspectives occurs as part of the meeting. Members of the group may mentally move into reflective thinking during parts of the meeting, engaging in reading parts of the report. Most of the effort is spent in reaching consensus.

Although the meeting occurs over a period of hours, the time-scale for organizational learning is significantly longer. If the organization is said to learn, then that learning is expressed in the behavior of the organization, which is visible as the changes in decisions over time. Clearly these changes are more apparent over a series of decisions rather than within a single meeting. Thus the time-scale for organizational learning is significantly longer than the hours during which a meeting takes place.

BEST COPY AVAILABLE

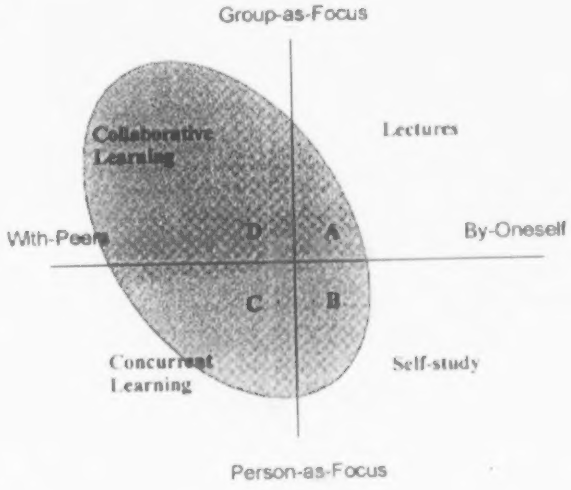


Figure 6. Business Meeting

Table of Contents

Conclusion

We believe that the debate between individual and collaborative learning is based on a false paradigm, that the two are on opposite ends of a spectrum. Instead, we have presented a model of individual and group processes in learning that reflect the singularity of the learning process as well as its orientation to the self or the group. From this model, we arrive at four learning categories, representing different mixes of process activity and orientation. They are: Lectures, Self-Study, Concurrent Learning, and Collaborative Learning.

These learning categories are not mutually exclusive or superior or inferior to one another. Rather, they represent different approaches to individual and group needs and roles, addressing different stages in the integration of information into knowledge for the learner as an individual, and as a member of a social group. In actual learning situations, individuals and groups can and do move between the processes described for each learning category. Specific patterns of these process mixes are evident in a number of examples of typical learning situations.

It is our belief that with this understanding, we can view the resource allocation and technology support issues of education and learning not as taking from Peter to pay Paul, but that these learning categories are all essential and inter-dependent. The questions should not be how we replace one by the other, but how we best use them most appropriately, both as strategies and as processes within learning situations, and ensure that learners are given the opportunity to benefit from each as needed.

A further benefit that we have gained from the model is that it has given us explicit guidelines to follow in the design of new learning products, both in who and what processes are being targeted, as well as what factors must be taken into account for them to be effective. This understanding spans the spectrum of learning systems as they currently exist, to new learning systems as they evolve, catalysed by developments in communications and information capabilities. What we hope to achieve from this understanding is a set of learning approaches and tools that span the entire Circle of Learning, and are integrated in ways which reflect the different processes and needs in each quadrant.

References

Bannon LJ. 1995. Issues in computer supported collaborative learning. In C. O'Malley, editor, *Computer Supported Collaborative Learning*, volume 128 of NATO ASI Series, Series F, Computer and Systems Sciences. Springer Verlag, Berlin.

Brown JS, Duguid P. 1991. Organizational learning and communities of practice. *Organization Science* 2(1):40-57.

Carley K. 1992. Organizational learning and personnel turnover. *Organization Science* 3(1):20-46.

Dillon JT. 1988. *Questioning and teaching: A manual of practice*. Teachers College Press: New York.

Felder R. 1995. We never said it would be easy. *Chemical Engineering Education* 29(1):32-33.

Fjuk A. 1995. Towards an analytical framework for CSC distance learning. In J. L. Schnase and E. L. Cunnius, editors, *Proceedings of CSCL '95: The First International Conference on Computer Support for Collaborative Learning.* Lawerence.

Garko MG, Kough C et al. 1994. Myths about student-faculty relationships: what do students really want? *Journal on Excellence in College Teaching* 5(2):51-65.

Gordon J. 1992. Performance technology: blueprint for the learning organization. *Training* 29(5):27-36.

Graesser AC, Person NK. 1994. Question asking during tutoring. *American Educational Research Journal* 31(1):104-137. Spring, 1994.

Johnson DW, Johnson RT, Smith KA. 1991. *Active learning: cooperation in the college classroom*. Interaction Book Company, Edina, Minnesota.

Johnson DW, Johnson RT. 1991. Cooperative learning: increasing college faculty instructional productivity. ASHE-ERIC Higher Education Report No. 4. George Washington University.

Kaye A. 1992. Learning together apart. In A. Kaye, editor, *Collaborative Learning Trough Computer Conferencing: The Najaden Papers*, volume 90 of NATO ASI Series, Series F. Springer-Verlag, Berlin.

Macedonia M, Brutzman D. 1994. Mbone provides audio and video across the internet. *Computer* 27(4):30-36. April 1994.

McArthy V. 1995. Desktop videoconferencing: still a rough cut. *Datamation* 41. May 15, 1995.

Mumford A. 1992. Individual and organizational learning: the pursuit of change. *Management Decision* 30(6):143-148.

Nunamaker J, Dennis AR, et al. 1991. Electronic Meeting Systems to Support Group Work, *Communications of the ACM*, 34(7):40-61.

Ransbury MK, Harris SA. 1994. Study abroad: the reality of building dynamic group learning. *Journal on Excellence in College Teaching* 5(2):97-107.

Rettinger LA. 1995. Desktop videoconferencing: technology and use for remote seminar delivery. Master's Thesis, Dept. Computer Engineering. North Carolina State University.

Senge PM. 1992. Building learning organizations. *Journal of Quality and Participation* 15(2):30-38.

Taylor S. 1992. Managing a learning environment. *Personnel Management* 24(10):54-57.

Turoff M, Hiltz SR. 1982. Computer support for group versus individual decision decisions. *IEEE Transaction on Man, Machine and Cybernetics*. 30(1):82-91.

Vetter R. 1995. Videoconferencing on the internet. *Computer* 28(1):77-79. January 1995.

Vowles A. 1993. Gaining competitive advantage through organizational learning. *CMA Magazine* 67(3):12-14.

Watson R, DeSanctis G, et al. 1988. Using a group decision support system to facilitate group consensus: Some intended and unintended consequences. *MIS Quarterly* 12(3):463-478.

Table of Contents

About the Authors

Ernest JH Chang, MD PhD

echang@axia.com

Ernie Chang is the Executive Vice-President of Axia Multimedia Corporation, which he founded in 1993 to commercialize innovations in interactive multimedia learning strategies that he created and patented, while at the Alberta Research Council, where he was Department Head of Advanced Computing & Engineering from 1987 to 1993. This Department become Canada's largest applied artificial intelligence group, and has performed more than $20 million of research, development and contract activity in all areas of artificial intelligence and robotics, including intelligent tutoring systems and expert systems.

At Axia, he was the Executive Producer of five consumer CD-ROMs, four in the Axia "Know Your Birds" series, and the highly rated "Know Your Combat Jets." Presently his responsibilities are in the areas of learning strategies and health care products.

His current research and technical interests are in technology-based learning strategies, cognitive technologies, and applications of computer-based interactive multimedia. He has a PhD in Computer Science from the University of Toronto, and an MD from UBC. Prior to joining the Alberta Research Council, he was an Associate Professor in the Department of Computer Science at the University of Victoria.

Don Simpson, PhD

Dr. Don Simpson is currently Chief Mentor of the AXIA Innovation Mentoring Unit (AIMU), a business unit of the AXIA Multimedia Corporation.

Simpson has been involved in innovative organizational development work in over 60 countries. Included in this work has been his involvement in the design, development, and operation of a number of breakthrough institutions concerned with social, economic, and education issues. He has worked at all levels of the education system and is Professor Emeritus at the University of Western Ontario in London, Ontario. In addition to his role as an educator, he has been a researcher, a consultant, and an entrepreneur who created and ran two consulting businesses, as well as founding and managing a number of nonprofit organizations. In 1991, he created the International Institute for Innovation (the Triple i), a global network which has the competence, flexibility, and entrepreneurial flair to assist a wide variety of organizations in Canada, Europe, and Asia to survive and thrive in these difficult times. The Mentoring Unit of the Triple i joined with AXIA Multimedia Corporation in January 1996.

Copyright 1997 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://olam.ed.asu.edu/epaa

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey@olam.ed.asu.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

John Covaleskie
Northern Michigan University

Andrew Coulson
a_coulson@msn.com

Alan Davis
University of Colorado, Denver

Sherman Dorn
University of South Florida

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
hmwkhelp@scott.net

Thomas F. Green
Syracuse University

Alison I. Griffith
York University

Arlen Gullickson
Western Michigan University

Ernest R. House
University of Colorado

Aimee Howley
Marshall University

Craig B. Howley
Appalachia Educational Laboratory

William Hunter
University of Calgary

Richard M. Jaeger
University of North Carolina--Greensboro

Daniel Kallós
Umeå University

Benjamin Levin
University of Manitoba

Thomas Mauhs-Pugh
Rocky Mountain College

Dewayne Matthews
Western Interstate Commission for Higher Education

William McInerney
Purdue University

Mary P. McKeown
Arizona Board of Regents

Les McLean
University of Toronto

Susan Bobbitt Nolen
University of Washington

Anne L. Pemberton
apembert@pen.k12.va.us

Hugh G. Petrie
SUNY Buffalo

Richard C. Richardson
Arizona State University

Anthony G. Rud Jr.
Purdue University

Dennis Sayers
University of California at Davis

Jay D. Scribner
University of Texas at Austin

Michael Scriven
scriven@aol.com

Robert E. Stake
University of Illinois--UC

Robert Stonehill
U.S. Department of Education

Robert T. Stout
Arizona State University

---

archives | abstracts | editors | board | submit | comment | subscribe | search

EPAA
- enter the archives
- browse the abstracts
- the editors
- the edit board
- submit article
- submit commentary
- search
- subscribe

volume: • 1 • 2 • 3 • 4 • 5 • 6 • 7

This article has been retrieved 1203 times since March 18, 1997.

Education Policy Analysis Archives

Volume 5 Number 8 **March 18, 1997** **ISSN 1068-2341**

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass Glass@ASU.EDU.

College of Education

Arizona State University,Tempe AZ 85287-2411

Copyright 1997, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

Response to Haskell: "Academic Freedom, Tenure, and Student Evaluation of Faculty"

Jeffrey E. Stake
School of Law-Bloomington
Indiana University

Abstract

Haskell (1997) argued that the administrative practice of student evaluation of faculty is a threat to academic freedom. However, before that claim can be substantiated, several prior questions must be addressed: To whom does academic freedom belong? Individual faculty? The academy? Whose actions can violate the right? Can any lines be drawn based on whether the substance or form of classroom behavior is influenced? And still another crucial point is whether a body can violate academic freedom without any intent to interfere with or control the substance of what is said to students.

Almost anything that can be done to undermine the administrative practice of getting students to evaluate teaching ought to be done. One of my major concerns is that the process of asking students their opinions undermines the trust and faith they need to place in the teacher. Instead of saying, "Here is a great scholar and teacher; learn from her what you can," the administration of evaluation forms says to students, "We hired these teachers, but we are not sure they can teach or have taught you enough. Please tell us whether we guessed right." As my father likes to say "The overexamined life is not worth living either." In this case, asking students for their opinions focuses the attention of students on the acting and special effects, rather than the message. I think students need to have trust in teachers to learn much from them. The evaluation forms undermine that trust.

I also believe that student evaluations can strongly influence the behavior of teachers, and for the worse. I changed my teaching dramatically because I was told by my Dean at the time that I had to keep the customers satisfied if I wanted to get tenure. (And I have not changed back since getting tenure.) I would not contend that the changes I made improved my teaching.

That said, I am afraid I have not been convinced by Haskell's arguments that the evaluations violate academic freedom. If *I* were to have my students fill out forms on my teaching, surely it would not violate my academic freedom. What if a colleague wishing the best for my success convinces me to do so? Does that violate academic freedom? If not, how about a well-meaning teaching committee? An avuncular Dean in a friendly tone, or in a threatening tone?

A closely connected question is whether academic freedom belongs to the academy or to individual teachers. I am unclear on this point and see arguments on both sides. Seen from one perspective, academic freedom is freedom for the academy to teach and research without control from outside, not for faculty members to be free from constraints imposed by the faculty or administration. When the academy imposes student evaluations on itself, there is no violation of academic freedom, however bad the teaching gets in response. Robert O'Neil, in his excellent book *Free Speech in the College Community* (Indiana University Press 1997), offers a small degree of support for this view:

> "Policies we impose on ourselves are ... much harder to challenge in court than are the policies government visits upon us."(p.189)

However, other passages in O'Neil's Chapter 8 convince me that he, at least, would probably not buy the proposition that academic freedom belongs to the university as an institution and not to the professoriate and professors. In his discussion of university attempts to limit research, O'Neil wrote (although without offering support) "If academic freedom means anything, it means that professors may speak out in institutionally embarrassing ways or in ways that may be at variance with institutional values and mission." (p. 178) This illustrates the viewpoint that academic freedom belongs to the professoriate, not the university. It is fair to say that its ownership is no simple matter, and resolving it one way or the other would not settle the question of the wisdom of using student evaluations of faculty.

The case that student evaluations violate academic freedom was not made to my satisfaction in the Haskell piece, in spite of the many other good points he has made against their use. Certainly the evaluations affect our classroom behavior, influencing both the style and content of our presentations. But that alone is not enough. As O'Neil concedes, academic freedom does not stop universities from imposing a large set of regulations on research.

> "On the one hand, researchers must and do accept all sorts of restrictions and conditions. The effect of some such constraints on the scope of inquiry is not trivial." (p. 176)

And even subject matter is not beyond control of the university. O'Neil makes the point that a geographer who teaches the earth is flat:

> "may forfeit the safeguards of academic freedom for flouting the very values on which a community of inquiry and scholarship depends". (p. xii)

But I would go further than that. Certainly I could properly be pulled from the classroom if I insisted on teaching only what everyone else would call "art history" in my "Property Law" course, even if I teach a stellar art history course. We cannot leave all choices of substance to individual teachers.

Haskell does not give us a way determining what actions violate academic freedom. He has left some of the most basic issues unresolved, indeed even unaddressed. Who owns the freedom and, conversely, whose actions can violate the right? Can any lines be drawn based on whether the substance or form of classroom behavior is influenced? And still another crucial point is whether a body can violate academic freedom without any intent to interfere with or control the substance of what is said to students. Similarly, does the faculty member have to be aware that the administrative (or other) action is influencing her behavior? It may be too much to ask for clear tests to be enunciated, but it is not too much to ask that these issues be addressed in some way.

So how do we draw the line as to what sorts of academic behavior administrators can control without infringing upon academic freedom? I have not yet found an answer. But those making the claim that student evaluation forms go too far could help their case by offering some way to draw that line. On the other hand, insisting on that asks for too much, for no one has yet accomplished the task.

I am not arguing that all line drawing and decision making should be done in a legislative manner. It is fine to say, in the style of common law judges, "this infringes academic freedom," without setting forth a set of rules for making similar decisions in the future. If that is the approach taken, however, at least some comparisons should be made to other, well-accepted and established, violations of academic freedom. Those comparisons might lead the writer to discuss why this particular bad decision, to have students evaluate their instructors, needs to be corrected from outside the academy by courts rather than by the academy itself, which seems to be the implication of the argument that such evaluations violate academic freedom.

References

Haskell, R.E. (1997). Academic Freedom,Tenure, and Student Evaluation of Faculty. *Education Policy Analysis Archives.* Vol. 5 No. 6 (Online: http://olam.ed.asu.edu/epaa/v5n6.html).

O'Neil, R. (1997). *Free Speech in the College Community.* Indiana University Press.

About the Author

Jeffrey Evans Stake

JeffStake@law.indiana.edu

Home page: http://www.law.indiana.edu/hyplan/jeffreye.html

- JD Georgetown, 1981
- BS Math and Psychology, U of Illinois, 1975
- Graduate work in Education, U of Illinois, 1976
- Clerked for Judge Oscar H. Davis, US Court of Claims 1981-2
- Worked as lawyer for Covington & Burling, Washington DC 1982-5
- Indiana University School of Law-Bloomington since 1985; Professor of Law since 1992

Copyright 1997 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://olam.ed.asu.edu/epaa

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey@olam.ed.asu.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

John Covaleskie
Northern Michigan University

Andrew Coulson
a_coulson@msn.com

Alan Davis
University of Colorado, Denver

Sherman Dorn
University of South Florida

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
hmwkhelp@scott.net

Thomas F. Green
Syracuse University

Alison I. Griffith
York University

Arlen Gullickson
Western Michigan University

Ernest R. House
University of Colorado

Aimee Howley
Marshall University

Craig B. Howley
Appalachia Educational Laboratory

William Hunter
University of Calgary

Richard M. Jaeger
University of North Carolina--Greensboro

Daniel Kallós
Umeå University

Benjamin Levin
University of Manitoba

Thomas Mauhs-Pugh
Rocky Mountain College

Dewayne Matthews
Western Interstate Commission for Higher Education

William McInerney
Purdue University

Mary P. McKeown
Arizona Board of Regents

Les McLean
University of Toronto

Susan Bobbitt Nolen
University of Washington

Anne L. Pemberton
apembert@pen.k12.va.us

Hugh G. Petrie
SUNY Buffalo

Richard C. Richardson
Arizona State University

Anthony G. Rud Jr.
Purdue University

Dennis Sayers
University of California at Davis

Jay D. Scribner
University of Texas at Austin

Michael Scriven
scriven@aol.com

Robert E. Stake
University of Illinois--UC

Robert Stonehill
U.S. Department of Education

Robert T. Stout
Arizona State University

archives | abstracts | editors | board | submit | comment | subscribe | search

EPAA Commentary

Contributed Commentary on
Volume 5 Number 6: Haskell *Academic Freedom, Tenure, and Student Evaluation of Faculty*

and

Volume 5 Number 8: Stake *Response to Haskell: Academic Freedom, Tenure, and Student Evaluation of Faculty*

---

21 May 1997

Michael Theall

On drawing reasonable conclusions about student ratings of instruction: a reply to Haskell and to Stake.

Kenneth Ebel was right when he said (1983, p. 65) "No corner of the university.lacks faculty members who fulminate against student evaluations, with little or no examination of the large body of research ...that underlies the practice." Lately, there has been a lot of discussion attempting to causally like student ratings to problems such as "grade inflation," and to suggest that ratings somehow violate the great traditions of academic life. I do not mean to trivialize these issues. They are important and deserve discussion. Rather, I am targeting the use (perhaps, misuse) of these issues in efforts to discredit or do away with faculty evaluation practices in general and student ratings in particular.

"Grade-inflation" (i.e., the rise in "average" grades since 1990 or is it 1950, or 1890, or the dawn of time) is always a hot topic because it's easy to make sweeping generalizations about "today's students," especially in comparison to some more-than-likely-mythical group of "better" students from some time past (especially a time "...when I was a student"). While it is probably true that a lot of things have changed about higher education including the standards used to grade, the expectations of both faculty and students, and the overall grade profile, it is very risky to point to one reason for these changes. There just isn't evidence for such a simplistic leap.

When coupled with the sometimes hysterical rhetoric about student evaluations, the inflation complaint itself gets inflated. "Ratings" become the cause of the downfall of higher education. They are, as one anonymous and angry professor wrote to me "...a corrupt practice of the 60's....A sop to students from administrators who are unwilling or unable to do anything to really improve teaching." More recently, in an article ripe with rhetoric and loaded with misinterpretations of the literature, Robert Haskell (1997) suggested that ratings are a violation of academic freedom (and of course, the cause of grade inflation). Haskell included a long quote from a chapter Jennifer Franklin and I (1990) did in our New Directions for Teaching and Learning issue #43 (Theall & Franklin, 1990b) While our point was that ratings practice must be improved, Haskell used the quote to supplement other citations as evidence in building his case that ratings were an essentially unreliable form of data that should be done away with ... a gross misinterpretation of our intent.

Frankly, I don't like to recommend articles like this to those not actively involved in ratings research or practice because such writings can mislead readers who aren't really familiar with the cited ratings literature. I'm sorry if that sounds elitist: it isn't intended to be, but I do have a reason for noting it. We (Franklin & Theall, 1989) found that ignorance of evaluation/measurement literature and methods correlated significantly with negative faculty opinions about students and student ratings. I note this because discussions about ratings are so often filled with misinformation. For example, Haskell, (in reference note #26) says that "...Cohen's (1983) review reinforces our earlier conclusion that student ratings are inaccurate indicators of student learning...". Peter Cohen's most recognized contribution to the ratings literature is the meta-analysis of multisection validity studies in which performance on a common final exam was correlated with ratings (1981). There was a .43 correlation between ratings and performance on those exams. In other words, sections with highly rated instructors had higher average scores which could not be attributed to differential grading standards or sampling errors. This is evidence for the validity of ratings, the exact opposite of what Haskell says. Calling that relationship grade inflation is simply ridiculous!

Academic freedom has been defined in many ways, but never before in a way that suggests the construct (tradition? principle? tenet?) is vulnerable to the influence of student ratings. The only comparable complaint I've heard came from a faculty member who missed classes frequently because he attended numerous religious services. He told me that since the ratings instrument contained an item about the instructor being available and his ratings on that item were depressed due to his frequent absences, the student ratings process violated his constitutionally guaranteed freedom of religion. Fortunately, in a reply to Haskell in a later issue of EPAA (V5#8), Jeffrey Stake (1997) does us the favor of providing a strong counter-argument to the academic freedom issue. Unfortunately, he then proceeds to propose that "Almost anything that can be done to undermine the administrative practice of getting students to evaluate teaching ought to be done" (p. 1). Stake says that ratings undermine the faith and trust students must place in teachers. Given his opinion of ratings, I assume that he must then feel that teachers have no need to place faith or trust in their students. This paternalism is antiquated and unrealistic. Stake also proposes that administrative use of ratings creates an image that institutions are asking if they "guessed right" about the faculty they hired. Drawing this in the other direction, I suppose that assigning grades creates the image that institutions are asking if they "guessed right" about the students they admitted. Finally, Stake proposes that asking students' opinions "... focuses the attention of students on the acting and special effects, rather than the message." (p. 2) I haven't seen many evaluation forms which contain items on either of these topics. Stake's reference is probably to "educational seduction," the skill of the infamous "Dr. Fox" who supposedly entertained students and received high ratings despite the fact that he delivered no content (Naftulin, Ware, & Donnelly, 1973). Many who do not care for ratings find one study that supports their position but ignore subsequent work (e.g. Perry et. al., 1979; Marsh & Ware, 1982) which points out problems with the original study and proceeds to clarify the issue. In this case, the determination that while style is no substitute for substance, "expressiveness" in a lecture can be a powerful aide to students' motivation and attention during lectures and thus, to recall and other aspects of eventual performance.

Jennifer Franklin and I did two studies which involved looking at grades and ratings: one specifically on grade inflation (1991) and one on the attributions of students about their academic performance (1990a). In the first study we looked at thousands of courses from one institution over a five year period during which mandatory ratings were initiated. We found a statistically significant but numerically minute increase in average grades but we also found that average ratings actually decreased in the same time period. So, even if faculty were trying to "buy" higher ratings by giving higher grades, students "weren't buyin' it."

In the attribution study we found that ratings of instructors were very consistent (i.e. reliable) across all grades. In other words, the ratings were essentially the same from 'A' students and 'F' students. This held up when class average ratings showed that the instructor was considered to be "among the best" as well as when the instructor was considered to be "average" or 'among the worst." There was no revenge for low grades or reward for high grades.

Major reviews (e.g., Marsh, 1987) have reported either near zero correlations between workload, difficulty, and ratings, or positive relationships indicating that more work and/or more difficult courses often receive higher ratings (e.g., Cashin, 1988). Greenwald & Gillmore (1996a, 1996b) however, report a positive grades-ratings relationship (i.e., higher grades - higher ratings) and a negative grades-workload relationship (i.e., higher grades - less work). However, they do not then propose a direct, negative ratings-to-workload relationship (i.e., less work- higher ratings) although they suggest it when they say (p. 14) "...finding a negative path between expected grade and workload is critically diagnostic of a causal effect of grading leniency." One might extrapolate that if less/easier work is related to higher grades, and higher grades are related to higher ratings, then less/easier work should also relate to higher ratings. Whether the syllogism holds remains a question. Even if it does, there is still a question about its implications for the student ratings process. Does a relationship - even a causal one, mean that ratings are inherently invalid or unusable? Greenwald and Gillmore note that ratings are imperfect but that flaws are "correctable" (1996b) and suggest making improvements rather than dismissing ratings. For the record, the relationship between grades and ratings in other studies is relatively consistent, with correlations normally between .3 and .4 (e.g., Feldman, 1976).

The counter-argument to the grades-ratings-inflation idea is that there is (and should be) a relationship between grades and ratings because good teaching begets good learning. Good learning results in both good grades and satisfied learners. This is still a compelling notion but Greenwald and Gillmore (1996A, 1996B) have clarified the definitions and controlled some of the variables more precisely. On the surface, their work suggests that there is a relationship which does not rely on the counter-argument and that when all else is held constant, giving higher grades will result in getting higher ratings. In other words, you can't account for the relationship solely on the basis of the "good teaching begets good learning" counter-argument.

One question about their findings is whether their finding is a psychometric, an instructional, or a psychological phenomenon. Consider that ratings are undeniably a measure of the satisfaction of learners with their learning experience perhaps more than they are a direct or absolute measure of the total quality of instruction. Though this sounds like heresy from a ratings proponent, it is a position that most researchers have held for some time. Consider also, that ratings are collected before students get their final grades and thus, their opinions must be based on expectations which are, in turn, based on performance to date. Ratings thus can not be said to reflect a disconfirmed expectancy about the overall course outcome (i.e., the course grade). This would only be the case if ratings were gathered after final grades were distributed and the final grade disconfirmed what was expected as a result of the experiences during the semester. So the ratings relationship is limited to experiences and results during the term rather than to the final grade.

So what could account for the grades-ratings relationship if not "better teaching begets better learning"? I propose the following.

Given that student satisfaction is related to the students' perceptions that they have received something of value in return for their tuition dollars, and given that a series of successful classroom and related experiences has unfolded during the semester, it is reasonable to assume that students will believe that they have learned something, will be satisfied, and will provide positive ratings. Now, in many classes (especially the lower level, undergraduate courses that populate many evaluation databases and by virtue of their enrollments and numbers, may disproportionately influence results) students are not the best judges of the breadth and depth of the instructor's knowledge or the extent to which the instructor has provided a complete, current, or even adequate treatment of the subject. If students have received good grades and positive feedback from their instructors they should: 1) be satisfied; 2) feel that they have learned something; and 3) be able to honestly rate their experiences and their instructors highly [even if the experience might be marginal using some other rating criterion coming from some other group]. The data they provide is thus both valid and reliable. Valid, because they are the appropriate providers of data and (we will assume for the moment) the instruments used ask at least face valid questions. Reliable because there will be little variance across responses (Marsh, 1987) and because their responses will not change much over time (Frey, 1976).

The question now becomes one which deals with: 1) the appropriateness of course content; 2) the standards used for grading; and 3) the question of whether lenient coverage and grading were deliberately chosen in order to influence ratings. The first two items can be and should be dealt with via curricular mechanisms such as departmental review of courses and content, and faculty agreement on standards for student work. The third item represents an ethical dimension that has much less to do with ratings than it does with behavior. If it is possible to manipulate students covertly (as in "dumbing down" courses so that students are led to believe that they are learning a lot when, in fact, they are only scratching the surface), then the problem belongs to the instructor and/or the department. Don't blame students or the ratings process. If the situation involves overt manipulation (as in making a 'ratings for grades' deal), then the problem belongs to the instructor, the students, the institution, and to higher education itself. Blame the perpetrators and those who let them get away with it, not the process of collecting, reporting, and using ratings data. Ultimately, I believe this suggests that ratings are reliable and valid indicators. If what they indicate doesn't please those who get and use the data, then the users, their departments, or their institutions should correct the situation, rather than discard the data or the processes of collecting, analyzing, or reporting it.

There are three points here. The first is that no system, however well constructed is safe from manipulation. The second is that there are probably legitimate instances of situations in which the established findings from the ratings literature are violated even though these situations are infrequent enough to influence the results of large database analyses which have been the bases of most research. The third is that ratings are a symptom, not a cause. Many opinions about ratings (like those of Haskell and Stake) seek to "blame" ratings for phenomena like grade inflation. Legitimate concern for the quality of teaching and learning should prompt us to examine the contexts of teaching and learning as well as the ratings that are provided by students, and to carefully review the other available kinds of data before jumping to erroneous conclusions about what "causes" what.

Robert Haskell suggests that we do away with ratings. Perhaps the weakest part of his article is what isn't there: constructive suggestions for improvement. The thrust is simply that we should do away with student ratings. One could just as well argue that we should do away with grades. Either way, no more nasty inflation! WRONG! There are appropriate reasons and uses for both. What's needed is more substantial research, less rhetoric, and better informed faculty and administrators who can distinguish between opinion and evidence, separate mythology from established findings, and explore the situational context in order to arrive at fair and equitable conclusions about faculty performance. Doing this won't violate anyone's rights. In fact, it should lead to better evaluation, and if data are carefully collected, reported, and used, better teaching as well.

REFERENCES

Cashin, W. E. (1988) Student ratings of teaching: a summary of research. (IDEA paper No. 20) Kansas State university, Division of Continuing Education. ERIC # ED 302 567

Cohen, P. A. (1983) Comment on a selective review of the validity of student ratings of teaching. Journal of higher Education, 54, 448-458.

Cohen, P.A. (1981) Student ratings of instruction and student achievement: a meta-analysis of multisection validity studies. Review of Educational research, 51, 281-309.

Ebel, K. (1983) The aims of college teaching. San Francisco: Jossey Bass.

Feldman, K. A. (1976) Grades and college students' evaluations of their courses and teachers. Research in higher Education, 4, 69-111.

Franklin, J. L. & Theall, M. (1991) Grade inflation and student ratings: a closer look. Paper presented at the 72nd annual meeting of the American Educational Research Association. Chicago: April 7. ERIC # ED 349 318

Franklin, J. L. & Theall, M. (1990) Communicating student ratings to decision makers: design for good practice. In M. Theall & J. Franklin (Eds.) Student ratings of instruction: issues for improving practice. New Directions for Teaching and Learning # 43. San Francisco: Jossey Bass.

Franklin, J. L. & Theall, M. (1989) Who reads ratings. Knowledge, attitudes, and practices of users of student ratings of instruction. Paper presented at the 70th annual meeting of the American Educational Research Association. San Francisco: March 31.

Frey, P. (1976) Validity of student instructional ratings. Does timing matter? Journal of Higher Education, 47 (3), 327-336.

Greenwald, A. G. & Gillmore, G. M. (1996a) No pain, no gain? The importance of measuring course workload in student ratings of instruction. Journal of Educational Psychology (in press). Available HTTP: http://weber.u.washington.edu/~agg/paingain.htm

Greenwald, A. G. & Gillmore, G. M. (1996b) Applying social psychology to reveal a major (but correctable) flaw in student evaluations of teaching Available HTTP: http://weber.u.washington.edu/~agg/asptramf.htm

Haskell, R. E. (1997) Academic freedom, tenure, and student evaluations of faculty: galloping polls in the 21st century. Educational Policy Analysis Archives V5 #6. Available HTTP: http://olam.ed.asu.edu/epaa/v5n6.html

Marsh, H. A. (1987) Students' evaluations of university teaching: research findings, methodological issues, and directions for future research. International Journal of Educational Research, 11, 253-388.

Marsh, H. W. & Ware, J. E. (1982) Effects of expressiveness, content coverage, and incentive on multidimensional student ratings scales. New interpretations of the Dr. Fox effect. Journal of Educational Psychology, 74, 126-134.

Natuflin, D. H. Ware, J. E., & Donnelly, F. A. (1973) The Dr. Fox lecture. A paradigm of educational seduction. Journal of Medical Education, 48, 630-635.

Perry, R. R., Abrami, P. A., & Leventhal, L. (1979) Educational seduction: the effect of instructor expressiveness and lecture content on student ratings and achievement. Journal of Educational Psychology, 71, 107-116.

Stake, J. E. (1997) Response to Haskell: academic freedom, tenure, and student evaluations of faculty. Educational Policy Analysis Archives V5 # 8. Available HTTP: http://olam.ed.asu.edu/epaa/v5n8.html

Theall, M. & Franklin, J. L. (1990a) Attributions or retributions: student ratings and the perceived causes of performance. Paper presented at the 71st annual meeting of the American Educational Research Association. Boston: April 16. ERIC #ED 319 764

Theall, M. & Franklin, J.L. (1990b) Student ratings of instruction: issues for improving practice. New Directions for Teaching and Learning # 43. San Francisco: Jossey Bass.

About the Author

Michael Theall, Ph.D.
Associate Professor & Director
Center for Teaching and Learning University of Illinois At Springfield
Springfield, IL 62794-9243

phone: 217-786-7157
fax: 217-786-7188
email: theall@uis.edu

Education

- BA St Anslem College (1963; English)
- MS Syracuse Univ. (1976; Instructional Technology)
- PhD Syracuse Univ. (1980; Instructional Design, Development, & Evaluation)

EPAA Commentary

Contributed Commentary on
Volume 5 Number 6: Haskell *Academic Freedom, Tenure, and Student Evaluation of Faculty*
and
Volume 5 Number 8: Stake *Response to Haskell's "Academic Freedom ... & Student Evaluation"*

On Michael Theall's (and implied et al.) "A Reply to Haskell and to Stake"

Robert E. Haskell

In "On Drawing Reasonable Conclusions about Student Ratings of Instruction: a Reply to Haskell and to Stake," Michael Theall (1997), Director of the Center for Teaching and Learning at the University of Illinois at Springfield, opens his critique of my article (Haskell, 1997a) by lamentably describing my piece as rhetorical. He then goes on to characterize my article: as an example of (1) faculty who "fulminate" against SEF , as (2) "simplistic," (3) "loaded with misinterpretations of the literature," as (4) "mythology," (5) exhibiting an "ignorance of evaluation/measurement literature," as containing (6) "sweeping generalizations," (7) "misinformation," as (8) "simply ridiculous!" (9)" ripe" with "hysterical rhetoric," as (10) assuming a "mythical group of better students" of some bygone era, asserting that (11) SEF "are the cause of grade inflation," as suggesting (12) "we do away with ratings." Continuing, Theall wrote that (13) "Perhaps the weakest part of his article is what isn't there: constructive suggestions for improvement," that I (14) "suggested that ratings are a violation of academic freedom," and finally---but not exhaustively---(15) that "Academic freedom has been defined in many ways, but never before in a way that suggests the construct (tradition? principle? tenet?) is vulnerable to the influence of student ratings." The last two items, at least, are correct. With this said, let me say I am perplexed by Theall's response.

I am perplexed for three reasons. First, because of what my article actually clearly said and did not say, second, because of Theall's own findings, and third because of the tone. Let me begin with what my article actually clearly said and did not say, and Theall's reaction to it. This will lead naturally into his own findings which, ironically, seem to support much of my thesis.

Theall's Inaccuracies and Misinterpretations

In anticipation of responses like Theall's, I clearly laid out the boundaries of my article quite carefully. Evidently, he either did not notice these boundaries or for some reason elected to ignore them. I will start with the more general accusations and proceed to the more specific, both of which , as I will document, are filled with inaccuracies. He says of my article, among other things that "Perhaps the weakest part of his article is what isn't there: constructive suggestions for improvement." I had indicated that I would not be addressing this issue, nor dealing in detail with findings on validity of SEF, or with grade inflation. I clearly indicated the boundaries: I said

> In explicating SEF , many closely related issues must be substantially bracketed. These related issues include (1) its validity (Cahn , 1987; Damron , 1996; Greenwald, 1996; Greenwald and Gillmore , 1996; Scriven , 1993; Seldin , 1984; Tagomori , Bishop, and Laurence. 1995), (2) the problem of defining teaching effectiveness, (3) general

> variables affecting SEF scores, (4) alternatives to SEF 's (5) classroom politically correct or popular standards and perceptions, (6) low student academic preparation, (7) age and gender discrimination issues (Feldman, 1983, 1993) (8) strategies for change, and (9) other integrally related issues such as their being largely responsible for lowered course standards, and grade inflation. Though these are important and related issues, they can only be addressed here in so far as they directly impact the focus on SEF and academic freedom.

Despite these clearly stated boundaries, and despite the fact of providing example citations to many of the areas like validity that I could not cover---I refer the reader to endnotes #3, 4, 5, 6 and 14 in my paper. Disregarding these limits, Theall proceeded to critique my article by concentrating on (a) validity and (b) grade inflation, exactly those areas that I said would have to be bracketed. I refused to be pulled into these two areas in my paper, and I refuse to be hauled into them now.

Both validity and grade inflation are too important to be briefly dealt with within the confines of this response by simply offering, as did Theall, a citation here and a citation there. Such an ad hoc approach to these important issues is little more than a statistical analogue of using verbal anecdotes to resolve the issue. In a more appropriate forum, if Theall would like to come up with something a little more systematic than the statistical anecdotes he offered, it might be worth responding. (Note 1) In any event, Theall refused to substantially address the main issue of my article: academic freedom and SEF .

I will now address some of the undocumented specific inaccuracies of Theall's critique. From numerous places in his response, he says,

> Robert Haskell suggests that we do away with ratings....The thrust is simply that we should do away with student ratings.... I am targeting the use (perhaps, misuse) of these issues in efforts to discredit or do away with faculty evaluation practices in general and student ratings in particular....building his case that ratings were an essentially unreliable form of data that should be done away with.

I did not say that SEF should be done away with. In fact in (a) my original article and (b) my response to Stake, (Haskell, 1997b) both of which Theall evidently read, I clearly stated:

> It is important to note at the outset, that it is not SEF per se that is the issue, but the impact of its use on salary, promotion, tenure decisions, and its impact on the delivery of quality education.

In addition in endnote #6 I said, again:

> If used correctly (see Copeland and Murry, 1996; Kemp and Kuman, 1990; Scriven, 1995, 1993, 1991; Seldin, 1984), SEF can be very useful instructionally, and when used in conjunction with other methodologically sound evaluation procedures and criteria, it can assist in informing an institution when a faculty does not pass muster as an effective teacher.

Let me make it clear once again: I am not against SEF , only their use for administrative purposes. As I said in my paper, while it is true that

> SEF may indeed validly describe incompetence...given (a) the conflicting data on their validity, (b) the way many institutions have constructed SEF instruments, (c) the often unsystematic statistical method by which SEF are interpreted, and especially (d) given

> the considerable weight accorded negative comments by only a few students in making tenure and promotion decisions, it would seem SEF can all to easily be used as a covert instrument for the elimination of tenure candidates and other faculty who may threaten student tuition dollars and perhaps ideological and popular culture agendas.

Note that I said "conflicting data on their validity," not that the literature demonstrates conclusively that they are completely invalid.

Given the number of instances on which Theall used the word "practical," I would think, given the time and resource intensive requirements for rendering SEF use valid, he would appreciate the fact that the likelihood of its appropriate use is not high. (In fact he does. See below). Theall , further claims that I make causal attributions from SEF to grade inflation. He wrote,

> Lately, there has been a lot of discussion attempting to causally link student ratings to problems such as grade inflation'.... Ratings" become the cause of the downfall of higher education.... Robert Haskell (1997) suggested that ratings are.... the cause of grade inflation)....it is very risky to point to one reason for these changes.

Nowhere in my article was it either suggested that SEF is "the cause" or was the term "cause" used or such an attribution made about SEF. I did suggest that SEF is

> largely responsible for lowered course standards, and grade inflation....are responsible for a considerable amount of grade inflation.

As scholars and rigorous researchers will likely agree there is a considerable difference between attributing a single causal connection and suggesting they contribute to grade inflation. I guess the statistical concept of variance was conveniently forgotten.

Theall claims that I "blame" students. He wrote, "Don't blame students." In anticipating such a reading of my article, I stated in the opening that to question SEF is often

> seen as not only a novel idea, but as an attack on either students, or a general attack on evaluating faculty.

I evidently anticipated this claim correctly. In further anticipation of such a reading of my article, in endnote #12 I said,

> It should go without saying, that not all students are the same. SEF vary by maturity, intellectual level, i.e., graduate student evaluations v. undergraduate.

Then later in the text I stated,

> Many students understand the above described ensuing consequences. A glance at articles from online student newspapers reveals strong sentiments against what many students consider the erosion of standards created by SEF . One student writer went so far as to say "We therefore suggest a boycott of the 1995 student/teacher evaluations. This boycott will provide a more effective means of communication than anything written on the evaluation itself. Something must be done about the trend of grade inflation. We as students refuse to contribute to the downfall of academia (Stern and Flynn, 1995) Some students are thus quite aware of the effects of SEF on their education.

Theall then goes on to claim that if there is grade inflation that one of the parties we should blame is faculty; that we should "Blame the perpetrators." There is a section in my paper entitled, "Faculty Complicity and Adaptation to SEF," in which I analyze this problem, but he does not mention any of the arguments I advance. I come to quite a different conclusion, however, than does Theall .

To my amazement, Theall simply notes, almost in passing, that I suggest that SEF "somehow violate the great traditions of academic life." What does he mean, somehow? The entire article was a fairly detailed "let me count the ways"--a listing, explaining, and accounting of just how I suggest SEF do violate academic freedom, both in the abstract and with concrete examples. So when he simply says of my article that "Academic freedom has been defined in many ways, but never before in a way that suggests the construct (tradition? principle? tenet?) is vulnerable to the influence of student ratings," I would suggest that Theall, tell us more about SEF and academic freedom, since this was the main thesis of the article.

I would now like to note very briefly and comment on some other claims made by Theall . My comments will be brief because, again, I dealt with them in my paper. First, he says that students "are the appropriate providers of data," mainly because they are paying "tuition dollars." I addressed these issues in my paper in discussing the misplaced (a) political metaphor of democracy applied to instruction, and (b) corporate metaphor of student as consumer.

Theall also claims that

> ratings are collected before students get their final grades and thus, their opinions must be based on expectations which are, in turn, based on performance to date. Ratings thus can not be said to reflect a disconfirmed expectancy about the overall course outcome (i.e., the course grade).This would only be the case if ratings were gathered after final grades were distributed and the final grade disconfirmed what was expected as a result of the experiences during the semester. So the ratings relationship is limited to experiences and results during the term rather than to the final grade.

In many if not most cases Theall's claim is incorrect. This is why: (a) In many, and maybe most, courses students receive test grades and the class curve/distribution (or at least should) throughout the semester, (b) the student grapevine is quite efficient is informing students what instructors grading practices are. So ratings thus can be said to reflect a disconfirmed expectancy about the overall course outcome.

Theall at least did make one technically valid set of points. Assuming SEF are on their face valid and reliable because there is "little variance across responses", says Theall ,

> The question now becomes one which deals with: 1) the appropriateness of course content; 2) the standards used for grading; and 3) the question of whether lenient coverage and grading were deliberately chosen in order to influence ratings. The first two items can be and should be dealt with via curricular mechanisms such as departmental review of courses and content, and faculty agreement on standards for student work.

Granting for the moment SEF validity and reliability (I probably would even really accept their reliability), I would like to address 1) and 2) in the above quote (I addressed the third above). On a collective level, I agree with Theall that faculty have not been responsible in developing and enforcing standards. In fact in endnote #34 I wrote:

> This is not intended as a blanket apologia for academia. There are many problems within the academy. In many other areas, I am a severe critic of my colleague's

collective behavior.

Theall's Work Supports Much of My Article

I noted at the beginning of this response that I was perplexed at much of Theall's reaction. I am particularly perplexed at his response to a passage I cited from his work. He says,

> Haskell included a long quote from a chapter Jennifer Franklin and I (1990) did in our New Directions for Teaching and Learning issue #43 (Theall & Franklin, 1990b) While our point was that ratings practice must be improved, Haskell used the quote to supplement other citations as evidence in building his case that ratings were an essentially unreliable form of data that should be done away with ... a gross misinterpretation of our intent.

I am perplexed for two reasons. First, it is really neither here nor there that the quote apparently does not reflect Theall's (et al.) intent. Whatever an author's intent was in using a quote does not necessarily reflect on other uses of the material, especially when the quote is not used to suggest an authors position.

Second, I am perplexed because much of Theall's chapter revolves around other material that he suggests demonstrate some of the severe problems in using SEF as they are currently used. The full quote to which Theall refers is as follows:

> Even given the inherently less than perfect nature of ratings data and the analytical inclinations of academics, the problem of unskilled users, making decisions based on invalid interpretations of ambiguous or frankly bad data, deserves attention. According to Thompson (1988, p. 217) "Bayes Theorem shows that anything close to an accurate interpretation of the results of imperfect predictors is very elusive at the intuitive level. Indeed, empirical studies have shown that persons unfamiliar with conditional probability are quite poor at doing so (that is, interpreting ratings results) unless the situation is quite simple." It seems likely that the combination of less than perfect data with less than perfect users could quickly yield completely unacceptable practices, unless safeguards were in place to insure that users knew how to recognize problems of validity and reliability, understood the inherent limitations of rating data and knew valid procedures for using ratings data in the contexts of summative and formative evaluation. (79-80).

The authors conclude by noting "It is hard to ignore the mounting anecdotal evidence of abuse. Our findings, and the evidence that ratings use is on the increase, taken together, suggest that ratings malpractice, causing harm to individual careers and undermining institutional goals, deserves our attention." (pp. 79-80).

Then in endnote #26 of my paper, Theall objects to my quoting Dowell and Neal noting that: "...Cohen's (1983) review reinforces our earlier conclusion that student ratings are inaccurate indicators of student learning..." (Note 2) Theall omits the rest of the Cohen's quote which said, "and therefore are best regarded as indices of consumer satisfaction' rather than teaching effectiveness." Even given that Theall (et al.) is talking about the logic in use of SEF, not its inherent validity, he goes on to say,

> Conversations with faculty and administrators...led increasingly to concerns about what users [e.g., chairmen; deans] were doing with the information we were providing. We saw that some departmental administrators, who routinely use ratings to make decisions about personnel, evaluation policy, and resource allocation, were not familiar

> enough with important ratings issues to make well informed decisions...Clearly stated disclaimers regarding the limitations of ratings data in particular circumstances appeared to have little effect on the inclination of some clients to use invalid or inadequate data...There are some fundamental concepts for using numbers in decision making. To the degree that these concepts are ignored, interpretations of data become, at best, projective tests reflecting what the user (e.g., a chairperson or dean) already knows, believes, or perceives in the data. Treating tables of numbers like inkblots ('ratings by Rorschach') will cause decisions to be subjective and liable to error or even litigation...

Theall (et al.) continue in that chapter to lay out carefully the methodological precondition for the reliable use of SEF . For example, in that chapter Theall's (et al.) wrote:

> Three types of errors come to mind immediately. The first involves interpretation of severely flawed data, with no recognition of the limitations imposed by problems in data collection, sampling, or analysis. This error can be compared to a Type I error in research -- wrongly rejecting the null hypothesis -- because it involves incorrectly interpreting the data and coming to an unwarranted conclusion. In this case, misinterpretation of statistics could lead to a decision favoring one instructor over another, when in fact the two instructors are not significantly different...(p.87-88)
>
> The second type of error occurs when, given adequate data, there is a failure to distinguish significant differences from insignificant differences. This error can be compared to a Type II error. -- failure to reject the null hypothesis - because the user does not realize that there is enough evidence to warrant a decision. In this case, failure to use data from available reports (assuming the reports to be complete, valid, reliable, and appropriate) may be prejudicial to an instructor whose performance has been outstanding but who, as a result of the error, is not appropriately rewarded or worse, is penalized. (p.87-88)
>
> The third type of error occurs when, given significant differences, there is a failure to account for or correctly identify the sources of differences. This error combines the other two types and is caused by misunderstanding of the influences of relevant and irrelevant variables. In this case, a personal predisposition toward teaching style.., may lead a user to attribute negative meanings to good ratings, or to misinterpret the results of an item as negative evidence when the item is actually irrelevant and there is no quantitative justification for such a decision. Any of these errors can render an interpretation entirely invalid.(p.87-88)
>
> Summarizing, Theall (et al.) wrote,
>
> Let us...state our goal in the following way: "The user will make decisions that are based on valid, reliable hypotheses about the meaning of data." In this case, the user should receive or construct working hypotheses that do the following things:
>
> - Take into account problems in measurement, sampling, or data collection and include any appropriate warnings or disclaimers regarding the suitability of the data for interpretation and use.
> - Do not attempt to account for differences between any results when they are statistically not significant (probably <.05).
> - Disregard any significant differences that are merely artifacts (for example, small

> differences observed in huge samples), which can technically be significant but are unimportant).
>
> - Account for any practically important, significant differences between results in terms of known, likely sources of systematic bias in ratings or reliably observed correlations, as well as in terms of relevant praxio logical constructs about teaching or instruction.
> - The user should also refrain from constructing or acting on hypotheses that do not meet these conditions...(pp. 87-89)...

Again, I said in my paper,

> SEF may indeed validly describe incompetence...given (a) the conflicting data on their validity, (b) the way many institutions have constructed SEF instruments, (c) the often unsystematic statistical method by which SEF are interpreted, and especially (d) given the considerable weight accorded negative comments by only a few students in making tenure and promotion decisions, it would seem SEF can all to easily be used as a covert instrument for the elimination of tenure candidates and other faculty who may threaten student tuition dollars and perhaps ideological and popular culture agendas.

So what is the problem of "intent" to which Theall refers?

Inappropriate Tone of Theall's Commentary

Theall's also makes brief mention of Stake's (1997) response to my article. While Stake can take care of himself, I would like to comment on one of Theall's comments as in my view it is indexical of Theall's setting up "straw men." Theall says,

> Stake's reference is probably to "educational seduction", the skill of the infamous "Dr. Fox" who supposedly entertained students and received high ratings despite the fact that he delivered no content (Naftulin, Ware, & Donnelly, 1973). Many who do not care for ratings find one study that supports their position but ignore subsequent work (e.g. Perry et al., 1979; Marsh & Ware, 1982) which points out problems with the original study and proceeds to clarify the issue.

The point is that Stake does not even come close to mentioning the Dr. Fox study, a study that used an actor to deliver lectures with nearly no content. Setting up such straw men, allows Theall to then claim to have demolished them. What the purpose of this was, I can not say.

Finally, in a critique which rather cavalierly takes on not only my article, but Stake (1997) and Greenwald, and Greenwald and Gilmore (1996), (Note 3) perhaps what is most disturbing is that Theall's attempt at rebuttal is, in tone, unjustifiably condescending and, in word, *ad hominem*. In referring to a research article of his (et al.), from which I quote in my article much to his dissatisfaction, he says,

> Frankly, I don't like to recommend articles like this to those not actively involved in ratings research or practice because such writings can mislead readers who aren't really familiar with the cited ratings literature. I'm sorry if that sounds elitist: it isn't intended to be but I do have a reason for noting it.

Theall should be more than "sorry;" as a scholar presumably knowledgeable in science, he should be embarrassed. He should also not flatter himself by suggesting his statement may be considered elitist; arrogant, perhaps, but not elitist. (Note 4) This is why: First, he has no legitimate reason to assume that I am not familiar with the literature (see below), just because I disagree with him and have reservations about SEF . His assumption seems to be based on the most fundamental of logical and statistical inference fallacies. Theall says,

> We (Franklin & Theall, 1989) found that ignorance of evaluation/measurement literature and methods correlated significantly with negative faculty opinions about students and student ratings. I note this because discussions about ratings are so often filled with misinformation.

Now as any first year college student learns in statistics, (a) correlation is not causation, and (b) one can not infer from a statistical generalization of a group, that an individual in that group is an indicator of that generalization. In fact this "logic" is the same as that used by people who exhibit prejudice: applying some generalization or central tendency characteristic of a specific group of people to an individual. Scriven (1988) makes this point in a similar context.

Second, were his claims not so clearly uninformed, they would be an insult not only to me, but to many other scholars as well. This is why that is: To say that scholars not actively involved in doing statistical or experimental research on a subject are not qualified to read, understand and comment on this research, immediately puts most theoretical physicists, theoretical biologist, and philosophers of science out of business. Hmmm. (Note 5)

References

Barnett, Larry, D. (1996). Are Teaching Evaluation Questionnaires Valid? Assessing the Evidence. Journal of Collective Negotiations in the Public Sector, 25, (4) 335-349

Cohen, P. A. (1983). Comment on a selective review of the validity of student ratings of teaching. Journal of Higher Education, 54, 448-458.

Franklin, J. L. & Theall , M. (1990). Communicating student ratings to decision makers: design for good practice. In M. Theall & J. Franklin (Eds .) Student ratings of instruction: issues for improving practice. New Directions for Teaching and Learning # 43. San Francisco: Jossey Bass.

Franklin, J. L. & Theall, M. (1989). Who reads ratings. Knowledge, attitudes, and practices of users of student ratings of instruction. Paper presented at the 70th annual meeting of the American Educational Research Association. San Francisco: March 31.

Greenwald, A. G. & Gillmore , G. M. (1997). No pain, no gain? The importance of measuring course workload in student ratings of instruction. Journal of Educational Psychology (in press).

Greenwald, A. G. & Gillmore , G. M. (1996). Applying social psychology to reveal a major (but correctable) flaw in student evaluations of teaching. University of Washington, Draft Manuscript, March 1.

Haskell, R. E. (1997b). Contributed Commentary on: Stake Response to Haskell: Academic Freedom, Tenure, and Student Evaluation of Faculty. Educational Policy Analysis Archives, 5 Available online: http://olam.ed.asu.edu/epaa/v5n8c1.html.

Haskell, R. E. (1997a). Academic freedom, tenure, and student evaluations of faculty: Galloping polls in the 21st century. Educational Policy Analysis Archives, 5. Available online: http://olam.ed.asu.edu/epaa/v5n6.html.

Natuflin , D. H. Ware, J. E., & Donnelly , F. A. (1973). The Dr. Fox lecture. A paradigm of educational seduction. Journal of Medical Education, 48, 630-635.

Scriven , M. (1988). The New Crisis in Teacher Evaluation: The Improper Use of 'Research-based' Indicators. Professional Personnel Evaluation News. (January) 1-8.

Stake, J. E. (1997). Response to Haskell: academic freedom, tenure, and student evaluations of faculty. Educational Policy Analysis Archives, 5. Available online: http://olam.ed.asu.edu/epaa/v5n8.html.

Theall , M. (1997). On Drawing Reasonable Conclusions about Student Ratings of Instruction: a Reply to Haskell and to Stake. Educational Policy Analysis Archives, 5. Available online: http://olam.ed.asu.edu/epaa/v5n8c2.html.

Thompson, G.E. (1988). Difficulties in interpreting course evaluations: Some Bayesian insights. Research in Higher Education, 28, 217-222.

Notes

1. Theall, suggested that I misinterpreted research on the SEF. For example, he says,

> Peter Cohen's most recognized contribution to the ratings literature is the meta-analysis of multisection validity studies in which performance on a common final exam was correlated with ratings (1981). There was a .43 correlation between ratings and performance on those exams. In other words, sections with highly rated instructors had higher average scores which could not be attributed to differential grading standards or sampling errors. This is evidence for the validity of ratings, the exact opposite of what Haskell says. Calling that relationship grade inflation is simply ridiculous!

Further, in at least partial defense before readers who may not be conversant with the literature, let me offer at least one somewhat lengthy statistical anecdote from Barnett (1996) that addresses Theall's belief in .43 correlations. In analyzing some of Cohen's (1981) data, Barnett explains,

> I turn now to a second statistical matter the interpretation of correlation coefficients. The numerical value of a correlation coefficient can be deceiving, even when it is 'statistically significant' (i.e., even when it is unlikely to have occurred by chance if no relationship exists). A correlation coefficient measures the degree to which change in the amount of the explanatory variable is accompanied by change in the amount of the effect variable, but the most beneficial feature of a coefficient is not its numerical value, which has no inherent, practical meaning. Rather, the square of the numerical value is the most advantageous aspect of a correlation coefficient, for the square indicates the proportion of variation in the effect variable that can be statistically attributed to variation in the explanatory variable. The research summary by Professor Cashin reported a correlation coefficient of .44 between student ratings of an instructor 'overall' and examination grades [8]. This coefficient means that 19.4 percent of the variation in student learning (as measured by course grades) is explained by variation in instructional quality (as measured by student ratings). If accurate, a correlation of this magnitude is 'practically useful,' as Professor Cashin said, though one must keep in mind that four-fifths of the variation in course grades remains unexplained and is attributable to other factors.

But does this correlation coefficient accurately estimate the relationship between student evaluations of teaching and student achievement? The best research on the magnitude of the relationship is the ' multisection validity study.' When it is ideally designed, such a study possesses the following features: each course included in the study has numerous sections; students are randomly assigned to sections; the sections of a course have different instructors but a common textbook and the same examination(s); all examinations for a course are constructed by a person who does not teach any section; and subjective (essay) components of examinations are graded by the person who developed them. A review of multisection validity studies cites one work that, the author of the review asserts, eliminates at least in part 'many of the criticisms of the multisection validity study' and ' provide[s] strong support for the validity of students' evaluations of teaching effectiveness' [9, p. 721]. However, the cited work which subjected the results of other multisection studies to a statistical analysis did not control a number of critical variables that could have generated or enlarged the relationship between student ratings of teachers and student achievement [10]. Among the missing variables that might have explained the relationship was the rigor of the requirements of the instructor (such as checks for student preparation and amount of material assigned), a factor that may vary considerably across sections of a single course. If the variable was related both to student ratings of instructional quality and to student achievement, a control for the variable could have markedly weakened or entirely eliminated the relationship originally found between student ratings and student achievement p.339

Another variable that the work omitted was the students' level of interest in the subject matter of the course prior to exposure to the teacher they later evaluated. As will be suggested below, neither of these variables should have been excluded from the analysis and left uncontrolled.

While the work did not incorporate a number of potentially important variables into its data analysis, the work is the source of a set of correlation coefficients (including the coefficient of .44) that Cashin suggested are credible estimates of the relationship between student ratings of teachers and student achievement. A reader of the reproduced coefficients can easily be misled, however, because Cashin failed to make clear that the coefficients may have been seriously confounded by variables whose influence was not removed. The failure to clarify this point is surprising inasmuch as Cashin explicitly stated that a control is necessary for one of the variables omitted by the work, namely, the interest students initially exhibited in the subject.' [12, p. 5]. P.340

Cohen treated student judgments of 'the amount and difficulty of the work the teacher expects of students' as one component of instructional quality. Whether it is an element of teaching quality or a separate factor, the amount and difficulty of work should be controlled under the conditions mentioned because it may explain much or all of the relationship detected between, on the one hand, evaluations of an instructor overall or on specific dimensions and, on the other, performance on examinations. From the studies he reviewed, Cohen calculated a negligible mean correlation coefficient for the relationship between the amount/difficulty of work and student achievement, but he also found a substantial range for the coefficients reported by the studies. Specifically, the interval for 95% of the coefficients extended from -.42 to +.39. Id, at 293, 295. Individual studies may thus involve a nontrivial association between the perceived difficulty of teachers and the examination performance of students (p.341).

I would also suggest that he reader see Greenwald's historical review of past studies and

Barnett's review of some of the most cited pieces that claim the validity of SEF.

2. I can understand how such quotes might be embarrassing (a) to an author who professes that SEF is basically if not completely valid instrument, and (b) to a practitioner trying to convince faculty and administrators to use them in faculty evaluation programs. I have never heard of a consultant who suggests that SEF have been shown to be invalid. Given Theall's critique let me make it very clear that I am not here indicting consultants. I have done and occasionally do consulting myself. I would like to say that Theall (et al.)---in word, I can not speak for deed---does recognize most of what needs to be done to make SEF use in administrative decisions reasonably ethical.

3. Requests for prepublication copies of Greenwald's articles should be sent to http://weber.u.washington.edu/~agg/

4. Theall goes on to say, "Stake says that ratings undermine the faith and trust students must place in teachers." Given his opinion of ratings, I assume that he must then feel that teachers have no need to place faith or trust in their students. This paternalism is antiquated and unrealistic. The logic by which it is unacceptable for Stake not to trust (valid?) perceptions of students, but acceptable for Theall not to trust the perceptions of scholars, escapes me, just a as does his calling Stake's view of students paternalistic and not seeing that his own views of his not trusting scholars to correctly read research is paternalistic. Though he does recognize his view sounds elitist, he says it is really not.

5. I wonder what Theall would say about child psychologists who have no children of their own, or about healthy psychotherapists who treat the mentally disordered.

EPAA Commentary

Contributed Commentary on Volume 5 Number 8: Stake *Response to Haskell: Academic Freedom, Tenure, and Student Evaluation of Faculty*

---

31 March 1997

Robert E. Haskell
University of New England

Stake seems to clearly agree that the administrative use of SEF should be eliminated, but remains curiously unconvinced by my arguments and multiple concrete examples suggesting that SEF violates academic freedom. He asks a number of questions regarding the conditions and circumstances of their use, and raises other significant questions regarding the meaning of academic freedom. I will address both these areas of concern, though I directly addressed many of them in the article to which Stake refers (Haskell, 1997).

In initially appraising Stake's response above, I specifically used the phrases "seems to clearly agree" and he remains "curiously unconvinced" for a reason. Stake opens his critique with,

> Almost anything that can be done to undermine the administrative practice of getting students to evaluate teaching ought to be done. One of my major concerns is that the process of asking students their opinion undermines the trust and faith they need to place in the teacher. Instead of saying, "Here is a great scholar and teacher; learn from her what you can," the administration of evaluation forms says to students, "We hired these teachers, but we are not sure they can teach or have taught you enough. Please tell us whether we guessed right...I also believe that student evaluations can strongly influence the behavior of teachers, and for the worse...[and]...I changed my teaching dramatically because I was told by my Dean at the time that I had to keep the customers satisfied if I wanted to get tenure....Certainly the evaluations affect our classroom behavior, influencing both the style and content of our presentations.

He then curiously goes on to write, "That said, I am afraid I have not been convinced by Haskell's arguments that the evaluations violate academic freedom." I am somewhat confused by this.

First, I am confused because it is not clear whether many of the objections Stake raises are specific to my paper or whether they are general questions; whether they belong to the realm of internal or external criticism. Second, I am confused because if he agrees that (a) SEF should be eliminated, (b) SEF undermine the trust that is apparently a prerequisite for student learning, (c) SEF can negatively influence faculty behavior, (d) he changed his own classroom behavior because of their influence on his candidacy to tenure, (e) because an administrator made it clear that his attaining tenure was in some measure dependent on achieving a positive rating students, and that (f) SEF can lead faculty to change the content of their courses, then (a) why does he remain unconvinced that SEF negatively impact academic freedom? And (b) if, indeed, SEF does not impinge on academic freedom, then why would he want to eliminate them?

Apparently--as indicated below--the answer to why he remains unconvinced that SEF impinge upon academic freedom could be that he is not sure where the boundaries of academic freedom lie. Stake then asks other questions. First, he asks,

> If I were to have my students fill out forms on my teaching, surely it would not violate my academic freedom.

My answer to this implied question was and is "no," for as I clearly indicated in my article,

> It is important to note at the outset, that it is not SEF per se that is the issue, but the impact of its use on salary, promotion, tenure decisions, and its impact on the delivery of quality education.

Second, to his question, "What if a colleague wishing the best for my success convinces me to do so? Does that violate academic freedom?" given my referring to the administrative use of SEF, not just any use of them, my answer is, accordingly, "No."

Third, to the related question, "how about a well-meaning teaching committee?" my answer would depend on the role of the committee. If the role of the committee was simply for confidential feedback to the individual faculty, given my above position stated in my article, my answer was and is again, "No." If, however, the committee is charged with policing teaching quality, and enforcing their views with some form of reward or withholding of reward, given my above position stated in my article, my answer was and is "Yes."

As for his fourth question regarding objecting to a suggestion to use SEF made by "An avuncular Dean in a friendly tone, or in a threatening tone?" my clearly implied answers are (for any familial-acting Dean) "maybe" and "yes." Maybe, depending on the context of the suggestion, and most certainly yes, if in a threatening tone (again, assuming their administrative use).

Moving on to the basic issue of academic freedom itself, Stake goes on:

> But I would go further than that. Certainly I could properly be pulled from the classroom if I insisted on teaching only what everyone else would call "art history" in my "Property Law" course, even if I teach a stellar art history course. We cannot leave all choices of substance to individual teachers.

I remain unclear as to what the problem is here regarding my paper. I said in footnote # 9

> While faculty are entitled to freedom of discussion and inquiry in their classroom, it is a generally recognized limitation that they should not introduce controversial matter which has no relation to their subject. What constitutes "controversial" and "no relation," however, often remains an open question.

Stake continues,

> A closely connected question is whether academic freedom belongs to the academy or to individual teachers. I am unclear on this point and see arguments on both sides....It is fair to say that its ownership is no simple matter, and resolving it one way or the other would not settle the question of the wisdom of using student evaluations of faculty.

First let me say, again, I am not sure whether Stake is unclear about the issue itself, or whether he is unclear regarding the issue of academic freedom in my paper. First, I said

> To further complicate matters the concept of academic freedom, like most abstract

terms is logically fuzzy around its edges.

Second, assuming Stake is unclear about the issue itself (the question of the ownership of academic freedom, notwithstanding for the moment), I would disagree that resolving academic freedom one way or the other would not settle the wisdom of using SEF. On my view, either way academic freedom is resolved SEF (used for administrative purposes) would violate academic freedom. I would perhaps agree with Stake if, and only if, a voluntary faculty policy assented to them (see below). I said in my paper,

> SEF can and do reflect these and other political and cultural conflicts, creating what the courts in other contexts have called a chilling effect on academic freedom....
>
> Academic freedom is a special concern of the First Amendment, which does not tolerate laws that cast a pall of orthodoxy over the classroom. While the two rights are not necessarily the same, they frequently and sufficiently overlap to trigger judicial scrutiny when faculty performance evaluation process threatens to impinge on the First Amendment (Copeland and Murry, 1996)

I also said in footnote # 29:

> For purposes of this paper I consider the mandated use or non use of SEF for salary, promotion and tenure by union contract agreements a special case of the "voluntary" nature of SEF.

Regarding whether academic freedom belongs (that is, to protect) individual faculty, I said,

> The university tenure function insures one of the only places in society where open dialogue on any issue no matter how unpopular or unorthodox can be critically examined without consideration of the political cost, without fear of reprisal, without the pressures of social taboos, social norms, faddish movements, personal notions of etiquette, and other immediate pragmatic pressures that exist in the culture in which it is embedded. Above all other roles, this is the defining feature of the university in a democratic society....Academic tenure, then, is not like seniority in business, civil service, or a union where the purpose of "tenure" is for the protection of the worker. Unlike these organizations, academic tenure is for the protection of the education function, not individual faculty jobs; it is for the protection of the role of the university in a democratic society.

Stake continues:

> But that alone is not enough. As O'Neil concedes, academic freedom does not stop universities from imposing a large set of regulations on research.

The apparently unresolved questions here are: Who is the university? Is it faculty policy? Is it independent administrative policy? Or is it a cooperative /compromise faculty/administrative policy? These issues appear to be open legal questions, or at least complex and conditional ones. Further Stake says,

> Haskell does not give us a way determining what actions violate academic freedom. He has left some of the most basic issues unresolved, indeed even unaddressed.....It may be too much to ask for clear tests to be enunciated, but it is not too much to ask that

> these issues be addressed in some way....So how do we draw the line as to what sorts of academic behavior administrators can control without infringing upon academic freedom? I have not yet found an answer. But those making the claim that student evaluation forms go too far could help their case by offering some way to draw that line. On the other hand, insisting on that asks for too much, for no one has yet accomplished the task. I am not arguing that all line drawing and decision-making should be done in a legislative manner. It is fine to say, in the style of common law judges, "this infringes academic freedom," without setting forth a set of rules for making similar decisions in the future. If that is the approach taken, however, at least some comparisons should be made to other, well-accepted and established, violations of academic freedom.

In my view, I find this confusing for two reasons. First, on a level, the entire paper addresses many of these issues. Second, I said

> In explicating SEF, many closely related issues must be substantially bracketed........the purpose of this paper is to coalesce arguments, and data, and explore implications of SEF.... this paper will review the issue in search of a more tutored policy....
>
> In a subsequent paper I will coalesce and deal more fully with legal ruling involving SEF in the denial of reappointment and tenure.

Stake further asks,

- Who owns the freedom and, conversely,
- whose actions can violate the right?
- Can any lines be drawn based on whether the substance or form of classroom behavior is influenced?
- And still another crucial point is whether a body can violate academic freedom without any intent to interfere with or control the substance of what is said to students.
- Similarly, does the faculty member have to be aware that the administrative (or other) action is influencing her behavior?
- Those comparisons might lead the writer to discuss why this particular bad decision, to have students evaluate their instructors, needs to be corrected from outside the academy by courts rather than by the academy itself, which seems to be the implication of the argument that such evaluations violate academic freedom.

These are all important questions, some of which I had planned to cover in two following papers. Some of his questions, however, I had not thought of and will subsequently consider.

References

Haskell, Robert E. (1997). Academic Freedom, Tenure, and Student Evaluation of Faculty: Galloping Polls in the 21st Century. Evaluation of Faculty" Education Policy Analysis Archives, 5 No. 6 (Online: http://olam.ed.asu.edu/epaa/v5n6.html).

Stake, Jeffrey E. (1997). Response to Haskell: "Academic Freedom, Tenure, and Student Evaluation of Faculty. Education Policy Analysis Archives, 5 No. 6 (Online:http://olam.ed.asu.edu/epaa/v5n8.html).

EPAA
• enter the archives • browse the abstracts • the editors • the edit board
• submit article • submit commentary • search • subscribe
volume: • 1 • 2 • 3 • 4 • 5 • 6 • 7

This article has been retrieved 795 times since March 21, 1997.

Education Policy Analysis Archives

Volume 5 Number 9 March 21, 1997 ISSN 1068-2341

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass Glass@ASU.EDU.

College of Education

Arizona State University,Tempe AZ 85287-2411

Copyright 1997, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

The History of the Reserve Officer Training Corps Among the Association of American Universities from 1982 to 1992: Review of Institutional Responses to ROTC Policy Regarding Homosexuals

Lee S. Duemer
Averett College

Abstract

This is a policy analysis, in a historical context, of how Association of American University institutions responded to Reserve Officer Training Corps policy excluding homosexuals. The time period for this study is 1982 to 1992. Qualitative methods are used to analyze data and arrive at conclusions. Secondary data provide additional depth and background. This study reveals seven different positions institutions have taken in response to ROTC policy, these include: supporting ROTC policy, neutrality, collective action, barring military recruiters from campus, distancing the institution from ROTC, and changing the campus climate. This includes examples taken from AAU institutions and rationales behind making policy decisions.

The purpose of this article is to develop a typology of institutional responses to Reserve Officer Training Corps (ROTC) policy regarding homosexuals, derived from the published responses of institutions composing the Association of American Universities (AAU). This will help in developing an understanding for those in higher education and the ROTC about a critical period in the history of this subject and how institutions responded to the conflict.

While much has been written about specific institutions and their responses to this issue, there has not been a comparative study examining the range of responses institutions have taken. This is important to higher education scholars, educational administrators and ROTC unit commanders, in order to develop a comparative understanding of how institutions responded to ROTC policy regarding homosexuals.

AAU institutions were selected as the focus of this study because of their prominence in American higher education. These institutions frequently encounter controversial issues before smaller colleges and universities, consequently, other institutions look to these flagship universities for guidance and instructions in how to deal with controversy when it develops.

The 1982 point of departure was chosen, as on January 28 of that year, Department of Defense (DoD) policy regarding homosexuals was revised in order to eliminate loopholes which allowed the admission and retention of homosexuals in the military. Under previous law, in existence since 1943, homosexual acts such as sodomy were considered illegal and punishable by imprisonment, however, whether the person was homosexual or not made no difference, the act was the focus of the law rather than the sexual identity of the individual (Berube, 1990). An individual caught in a homosexual act could avoid removal or imprisonment by claiming the act was an aberration, that they were not actually homosexual. In general, a homosexual was not subject to removal from the military so long as that person did not engage in homosexual acts.

The 1982 law eliminated this loophole so that simply admitting homosexuality, apart from homosexual activity, was ground for removal. The sexual identity of the individual, regardless of their actual behavior became the focus of the law. The year 1992 is chosen as the closure for this study because of the Don t ask, don t tell compromise developed that year by the Clinton administration.

Under the 1982 policy, homosexuals were prohibited from joining or serving in any branch of the military. This included ROTC branches of the Army, Navy and Air Force. According to DoD policy, a ROTC cadet could be removed from the Corps for engaging in, attempting to engage in, or soliciting another member to engage in homosexual acts; for stating one is homosexual or bisexual; or for marrying or attempting to marry one of the same sex. This process of removal was referred to as "disenrollment" (Clark, 1990; Gross, 1990). Because this was the official DoD policy, its enforcement in all branches of the armed forces and their ROTC units was mandatory.

This policy was in direct conflict with many institutional non-discrimination statements, which prohibited discrimination on the basis of sexual orientation. Consequently, the ROTC partnership with higher education became a source of friction on university campuses across the country. Institutions forming the AAU responded in a range of means, this included openly supporting ROTC, neutrality, collective action to compel the DoD to change its policy, banning military recruiters, organizing the removal of ROTC from campus, distancing the institution from ROTC, and finally, changing the university environment.

ROTC Policy Regarding Homosexuals from 1982 to 1992

Department of Defense policy, formalized in Directive 1332.14 Section (1) (H), banned homosexuals from serving in the armed forces. The policy on homosexuality stated:

> Homosexuality is incompatible with military service. The presence in the military environment of persons who engage in homosexual conduct or who, by their statements, demonstrate a propensity to engage in homosexual conduct, seriously impairs the accomplishment of the military mission. The presence of such members adversely affects the ability of the military services to maintain discipline, good order, and morale; to foster the mutual trust and confidence among the servicemembers; to ensure the integrity of the systems of rank and command; to facilitate assignment and worldwide deployment of servicemembers who frequently must live and work under close conditions affording minimal privacy; to recruit and retain members to the Military Services; to maintain the

public acceptability of military service; and to prevent breaches of security.

ROTC cadets who revealed that they committed homosexual acts, attempted to commit homosexual acts, or stated that they intended or desired to commit homosexual acts, or were discovered to be homosexual were discharged from the ROTC and technically required to reimburse the ROTC for cost of their education, although this was rarely done (Kosova, 1990).

Institutional Responses to ROTC Policy

Supportive Responses

Some administrators saw no conflict between the exclusionary policies of the ROTC and institutional non-discrimination policies (J.G. Kryway, personal communication, August 28, 1990; R.E. Jallette, personal communication, September 4, 1990). In some instances these supportive positions were publicly expressed, however, in other cases administrators have publicly responded negatively to the exclusion of homosexuals while privately affirming support for ROTC to their unit commander (R.E. Jallette, personal communication September 4, 1990; J.G. Kryway, personal communication, August 28, 1990; J.J. Petrick, personal communication, August 29, 1990).

One example of this was at Indiana University. In August 1990, a new Code of Ethics went into effect, which included a Sexual Orientation Clause, prohibiting discrimination on the basis of sexual orientation (J.G Kryway, personal communication, August 28, 1990). The president of the university stated that he supported the new institutional Code of Ethics but saw no conflict between it and ROTC practices (J.G. Kryway, personal communication, August 28, 1990).

An August 1990 memorandum from the ROTC unit commander at Johns Hopkins University stated that the president of the university advocated the continued presence of ROTC on campus (R.E. Jalette, personal communication, August 4, 1990). This statement was made in spite of the fact that ROTC and the university had conflicting policies with respect to institutional policy and sexual orientation. The president informed his unit commander that while he may be required by the board of trustees to take action requesting the DoD to change their policy, he "can live with conflict" (R.E. Jalette, personal communication, August 4, 1990). This was also the case when in May 1990, the Board of Regents at the University of Minnesota unanimously voted to require the university president, Nils Hasselmo, to write to the Secretary of Defense requesting that the existing military policy be changed. The President privately informed the ROTC Commander and Professors of Military Science that he would continue to abide by the existing DoD policy (M.D. Trout, personal communication, August 30, 1990).

To justify their support for ROTC, administrators cited its benefits to the nation. Seventy percent of the officers in the Army received their education at America's colleges and universities courtesy of the ROTC, including former Chairman of the Joint Chiefs of Staff, General Colin Powell, who received his undergraduate education through the ROTC at the City Colleges of New York (Shelton, 1985; Card and Elder, 1989; Kosova, 1990). This was and continues to be a significant point in light of the fact that this percentage is expected to remain unchanged if not expand as the government has made no efforts to increase the size of the military academies or promote large numbers of enlisted personnel into the officer corps (Malpass, 1985).

As competition for students increased, administrators also indicated the importance of a close partnership between ROTC and higher education in order to attract capable students (Malpass, 1985; Jaschik, 1993). While there were regional variances in the rate of enrollment decline, an effective ROTC program could help alleviate the problem (Malpass, 1985; Jaschik, 1993). In 1990 ROTC brought approximately $2 million to the University of Wisconsin at Madison in the form of scholarships and salaries (Kosova, 1990).

It is important, however, to point out that these positions were formulated without primary regard for the value of the program in its own right, rather, its excellence was derived from the fact that it served as a means to an end. This end was financial, in the form of academically qualified students with scholarships, and a positive relationship with the government in order to maintain and attract funding for research (Malpass, 1985; Kosova, 1990).

At the University of Kansas, Chancellor Gene Budig refused to implement a faculty resolution addressing the issue of ROTC excluding homosexuals (Swartz, 1990). In May 1990, the University Senate, composed of 52 elected faculty, students and staff, passed a resolution prohibiting the ROTC from conducting ROTC officer commissioning ceremonies on university property or involving university personnel in those ceremonies (Fagan, 1990; Swartz, 1990). ROTC commissioning ceremonies were traditionally held every spring before commencement (Swartz, 1990). However, the Chancellor, also a General in the Air National Guard, rejected the resolution (Swartz, 1990). His approval would have been necessary to make it university policy (Fagan, 1990). Neutrality

Institutions also responded neutrally to the issue of the ROTC and its policy of exclusion. Among the reasons cited were avoidance to involve the institution in a purely political issue, the benefits of ROTC to the institution with respect to student enrollment, and reluctance to interfere with positive and profitable government relations (Malpass, 1985; Trow, 1987; Kosova, 1990; Jaschik, 1993).

This position was largely grounded in the notion that it was improper to take a moral position on what may be a strictly political issue in which people can disagree in a moral forum (Trow, 1987). By transforming a political issue into a moral one and then taking sides while subsequently asserting the moral superiority of that side, it would become difficult to subordinate one's self to the common interest, the advancement and welfare of the institution (Trow, 1987). This type of political interference is what the former President of Harvard University, Derek Bok, referred to when he insisted that "universities have neither the mandate nor the competence to administer foreign policy, set our social and economic priorities, enforce standards of conduct in the society, or carry out other social functions apart from learning and discovery" (Trow, 1987).

This statement was paralleled by D. Bruce Johnstone, Chancellor of the SUNY system, who affirmed that in order for members of the higher education community to benefit from freedom of political interference, the price they have to pay is for institutions themselves not to become involved in political issues (Blumenstyk, 1991). This position argued that it is the politically neutral atmosphere of the university, which attracts people of diverse and varying points of view without the fear that their ideas or beliefs will be unfairly attacked (Trow, 1987; Blumenstyk, 1991). A political stance would deny the right of people with diverse political values to come together in a non-political environment to pursue other interests together, such as education, without regard to political differences (Trow, 1987).

Another reason some university administrators remained neutral is that taking the initiative with the ROTC put universities in a very uncomfortable position (Jaschik, 1993). According to the Assistant Chancellor for Legal Affairs at the University of California-Berkeley, Michael R. Smith, they would much rather have waited for the courts to address the controversy than for them to have had to take the awkward position of informing the DoD that it was wrong in its exclusionary policy (Fields, 1984b).

This was especially an issue for administrators at AAU institutions, who were hesitant to criticize the ROTC because of the millions of dollars in research grants that could have been placed at risk from negative relations with the DoD (Kosova, 1990). For administrators at the nation's major research institutions, this represented a financial decision, in which removal of the ROTC from a campus could have had a detrimental influence on securing federal funding for lucrative defense research (Kosova, 1990).

Collective Action

Significant in initiating collective action were two Pentagon research reports prepared by the Personnel Security Research and Education Center (PERSEREC) in Monterey, California. Still classified, draft copies of these were released to members of the House of Representatives Armed Services Committee. The reports found no correlation between homosexuality and evidence of security risk suitability, no evidence that homosexuals were a disruptive element in the armed forces, and finally urged their retention and expressed that the armed forces should consider admitting them (University of Pittsburgh ROTC Report, 1990). They reported that "homosexual men and women as a group are not different from heterosexual men and women in regard to adjustment criteria or job performance" (Card and Elder, 1989). The conclusions suggested that the armed forces should consider treating homosexuals as a non-ethnic minority as opposed to a form of deviancy or as criminals and recommended integrating open homosexuals into the military based on how African-Americans were integrated under Harry S. Truman (University of Pittsburgh ROTC Report, 1990).

In April 1990, less than one year after the release of the drafts of the two PERSEREC Reports, the Secretary of Defense, Dick Cheney received a letter from John M. Deutch, Provost of the Massachusetts Institute of Technology (Clark, 1990; Gross, 1990; Michaud, 1990). The letter by Mr. Deutch, a long-time Pentagon advisor, was a summary of points made by academics, highlighting the contradiction between the university principle of non- discrimination on the basis of sexual orientation and the presence of an ROTC which discriminates, adding that the presence of an ROTC which does discriminate cannot exist indefinitely (Gross, 1990; Maca, 1990). Though he made no specific details or deadlines, Mr. Deutch stated that because of ROTC policy, the risk was present that colleges and universities could withdraw from ROTC (Clark, 1990; Gross, 1990). He also cautioned Secretary Cheney that failure to reverse the ban would make it easy for dedicated opponents of the defense establishment to further their cause during a time of increasing calls for cutbacks in defense appropriations (Gross, 1990).

At the same time, an ROTC Advisory Committee at MIT was looking into the matter and arrived at the conclusion that it would require a long term commitment with collective action, and a time frame of three to fours years to phase out ROTC without harming students already enrolled in the program (Maca, 1990). Under this position, collective action would take place with the four higher education associations, the American Council on Education, the Association of American Universities, the American Association of State Colleges and Universities, and the National Association of State Universities and Land Grant Colleges (Maca, 1990).

The President of MIT, writing to the Undergraduate Association stated that:

> Both the Provost and I are troubled by the contradiction between MIT's policy of non-discrimination and the ROTC policy of discrimination on the basis of sexual preference, and we believe this ROTC policy should change...I believe that the military services not only should, but will, change their policies regarding sexual preference and ROTC programs, because the continuation of ROTC programs at leading colleges and universities will be greatly influenced by this question (Maca, 1990).

The MIT Dean of Undergraduate Education, overseeing the ROTC program, found this policy difference between the DoD and MIT to be "deeply troubling" (Maca, 1990).

At the same time as the efforts to address this problem at MIT, Kenneth A. Shaw, President of the University of Wisconsin System, introduced a resolution at an April, 1990 Association of American Universities meeting, opposing ROTC policy regarding homosexuals (Collinson, 1990b). The following month, the presidents of the four groups representing nearly all the nations colleges and universities hand delivered a letter to the DoD, which expressed serious concern and urged Secretary Cheney to change the DoD policy of barring homosexuals from the ROTC (Clark, 1990; Michaud, 1990). The letter stated that this policy was at odds with university affirmative action programs which prohibited discrimination on the basis of sexual orientation (Collinson, 1990b).

The letter was prepared and signed by Robert H. Atwell of the American Council on Education, Robert M. Rosenzweig of the Association of American Universities, Allan W. Ostar of the American

Association of State Colleges and Universities and Robert L. Clodius of the National Association of State Universities and Land Grant Colleges (Clark, 1990). The letter stated that the policy on homosexuals dated from the days of discrimination against blacks and other minorities. The letter further questioned the "curious anomaly" of discrimination based on sexual orientation, stating that "sexual orientation is the only basis in which discrimination is condoned within ROTC" and that other forms of "denial of equal opportunity on the grounds of race, gender, religion, nationality or political affiliation (have) long since been barred" (Collinson, 1990b). The letter stated that if this policy continued, the ROTC would lose not only the leadership of qualified homosexual cadets but also the support of the population who find discrimination intolerable in any form (Collinson, 1990b). The four presidents also requested a meeting with Secretary Cheney to discuss the issue (Michaud, 1990).

In late June of 1990, they received a reply from Secretary Cheney's office notifying them that "we do not plan to reassess the Defense Department's policy" and "that a meeting with the Secretary to discuss the issue would not be productive at this time" (Clark, 1990; Dodge, 1990). The letter also stated that there was no reason for the DoD to review the policy as up to that time they had won every court case on the issue (Michlaud, 1990).

This led to a series of letters written by AAU institution heads to the Secretary of Defense. In May 1990, the president of the University of Pennsylvania sent a letter to the Secretary of Defense, urging a change in policy concerning homosexuals (R.C. Miner, personal communication, August 30, 1990). This letter was written in response to a University Council resolution adopted that same month, which stipulated the suspension of the ROTC program by June of 1993 unless the DoD changed its policy (R.C. Miner, personal communication, August 30, 1990). In June of 1990, the President of the University of Oregon sent a letter to Secretary Cheney and Oregon U.S. Representatives seeking support for "changing the military's policy of discrimination" (Smith, 1990). In 1990, the presidents of Rutgers University and Indiana University also joined in calling formally with other AAU institution presidents for an end to ROTC policy (J.G. Kryway, personal communication, August 28, 1990; B.R. Maca, personal communication, September 4, 1990). At the University of Minnesota, the Board of Regents passed a resolution in May 1990, stating that ROTC was at odds with the university policy of non-discrimination and urged the President, Nils Hasselmo, to lobby Congress to change the policy (Collinson, 1990b; McNaron, 1991).

In June 1990, the President of Washington State University, who also served as a member of the Army ROTC Advisory Panel on ROTC Affairs, requested that Secretary Cheney modify the existing DoD policy (C.L. Pullman, personal communication, August 31, 1990). Mr. Smith considered the existing ROTC policy to be in direct conflict with a State Executive Order for the State of Washington, which mandated an end to discrimination against any class of individuals, including homosexuals (C.L. Pullman, personal communication, August 31, 1990).

In the fall of 1990, at the University of Pittsburgh, a committee was convened to collect information about the status of ROTC on campus (University of Pittsburgh ROTC Report, 1991). In February 1992, Chancellor J. Dennis O Conner accepted the eight recommendations, which the committee made,, with the exception of the discontinuation of the ROTC programs by 1997 should the DoD not revise its policy. Chancellor O'Conner felt that AAU universities should act as a group in removing ROTC programs, rather than each institution acting individually.

Banning Military Recruiters

Some institutions have reacted to ROTC policy by banning armed forces recruiters. These measures were taken not only in response to ROTC, but also to the overall military policy of excluding homosexuals. In this response, the laws which affect it and threats of repercussions from the DoD, are similar and in some cases the same as those affecting the ROTC (Norris, 1982).

After banning military recruiters, law schools at the UC-Los Angeles, Columbia University, Harvard University, New York University and Yale University, received letters from the Judge

Advocate General of the Army, Hugh Clausen, in May 1982, warning them that they could lose DoD funding if they continued to prohibit Army recruitment (Norris, 1982). Each of these schools has policies in effect, which prohibit recruitment by employers on campus that discriminate among a variety of ways, including sexual orientation. Each of these law schools also claimed that this measure was taken to ensure that their actions were consistent with their own internal policies.

The letters included the threat that if Army recruiters continued to be barred, the Judge Advocate General would recommend to the Secretary of Defense that no Army officers be trained at these institutions, and that "no Defense Department contracts be awarded to your university as long as our officers are denied the ability to recruit on campus" (Fields, 1982). The letters cited 1973 Public Law 92-436, Section 606, which prohibits the expenditure of DoD funds to any institution which bars military recruiters from its campus (Fields, 1984a,b). This law jeopardized the future support of tuition assistance for military personnel enrolled in the institution, research done for the military, and existing ROTC programs (Fields; 1982; Fields, 1984a; Card and Elder, 1989).

Initially, institutions were uncertain whether financial penalization applied only to the subordinate element of the institution, such as the school of law, or to the entire institution (Fields, 1982; Norris, 1982). A cutoff of DoD contracts was the most serious concern, collectively the five institutions received over $40 million in 1981 in military contracts (Fields, 1982; Norris, 1982). Later clarification made it clear that the law only applied to the subordinate elements (Fields, 1984a,b).

The schools stated that they would be willing to allow the military to recruit on campus if recruiters would sign a statement which all other employment recruiters sign. This statement called on recruiters to pledge that they did not discriminate in their hiring practices on the basis of sex, race, age, national origin or sexual preference (Fields, 1984,b). Armed forces recruiters refused to sign the non-discrimination statement or discuss the issue of homosexuals, none of the institutions modified their policies barring military recruiters and the threat of cutting military funding was never carried out (Norris, 1982; Fields, 1984a).

In May 1984, Gen. Clausen sent a second round of letters to the UC-Berkeley, UC-Los Angeles, Columbia University, Harvard University, University of Minnesota, New York University, University of Pennsylvania and Yale University (Fields, 1984a). These letters reiterated the earlier warning of financial repercussions regarding Public Law 92-436, Section 606 (Fields, 1984a). This second series of letters was sent as a follow-up, in order to inquire about the policies of these institutions regarding military recruiters and their law schools (Fields, 1984a; Fields, 1984b).

As before, institutions offered military recruiters the opportunity to recruit, provided they signed non-discrimination statements (Fields, 1984a; Fields, 1984b). Military recruiters did not sign the statements and no sanctions were taken against the institutions as the DoD did "not wish to engage in a confrontation with institutions of higher learning over their career placement policies" (Fields, 1984b).

In October 1989, the University of Iowa, College of Law joined the above institutions by banning military recruiters of the Judge Advocate General from conducting interviews in the building (Kosova, 1990; Roberts, 1990). In response to this, the ROTC unit on campus allowed the recruiters to use their facilities for recruitment purposes (Kosova, 1990; Roberts, 1990).

On September 20, 1991, the Office of Lesbian and Gay Concerns, part of the State Division of Human Rights in New York, ordered the law school of the State University at Buffalo to bar military recruiters from its campus citing that the armed forces discrimination against homosexuals conflicted with 1983 and 1987 Executive Orders issued by Governor Mario Cuomo outlawing discrimination based on sexual orientation (Gogola, 1991). The agency ruled that by allowing the recruiters on the 64 campus SUNY system the law school was violating its own policies prohibiting discrimination on the basis of sexual orientation (Mercer, 1992).

However, by the end of that same day, Governor Cuomo declared the Division of Human Rights ruling to be "unenforceable" because of a state education law which allows the military the same right to recruit on campus and that it was not within the authority of an executive order to affect the military (Gogola, 1991; Mercer, 1992). While the Division of Human Rights could challenge cases in state courts over alleged violations of New York law, they had no authority to challenge executive orders (Blumenstyk, 1991). This left the Division of Human Rights with no authority to enforce its ruling.

Administrators who sought a change in the DoD policy toward homosexuals did not consider this to be a major defeat in their efforts (Mercer, 1992). They did concede, however, that due to the size of the SUNY system, had the barring of recruiters held up, it would have been a major victory (Mercer, 1992). The Chancellor of the University of Buffalo stated that he would continue to "push to see that the military opens its doors" (Mercer, 1992).

In November 1991 a bill was approved in the Illinois State Legislature which would have prevented the state's colleges and universities from barring military recruiters or the ROTC (Blumenstyk, 1991; Gogola, 1991). The bill was designed to see that attempts by institutions across the country to bar recruiters were not successful in Illinois (Blumenstyk, 1991; Gogola, 1991). However, Governor Jim Edgar vetoed the bill, explaining that he felt it was an intrusion into the rights of higher education governing boards to regulate their own institutions (Blumenstyk, 1991; Gogola, 1991).

Removal of ROTC from Campus

The complete removal of ROTC from an institution, and severance of relations with ROTC, is another position administrators have considered (Jaschik, 1993). The appropriateness of severing the institution from ROTC has been derived from that fact that it is the local unit, which is in violation of institutional policy (Card and Elder, 1989). The object of putting pressure on the local unit has been to persuade the DoD at its highest authority to change its policy regarding homosexuals (Card and Elder, 1989). However, the idea of barring ROTC from campuses, in order to bring the institution into compliance with institutional policies prohibiting discrimination based on sexual orientation, is complicated by the responsibilities of land grant institutions to offer military training and the political repercussions of eliminating the ROTC at AAU universities (Collison, 1989; Kosova, 1989; Card and Elder, 1990).

At the University of Wisconsin-Madison, faculty members asked for the removal of ROTC from the campus, only to be overruled by the president and board of trustees (Michaud, 1990). In December 1989, at the first full meeting since 1970, faculty voted to urge the Board of Regents to remove the ROTC from campus and sever all university contacts by 1993 unless it changed its policy of excluding homosexuals (Card and Elder, 1989; Gross, 1990; Kosova, 1990). The faculty found ROTC to be in violation of institutional policies in addition to state law (Card and Elder, 1989). State law became an issue when state legislators passed Wisconsin Assembly Bill 70, which broadened non-discrimination laws to include sexual orientation (J.S. Riley, personal communication, August 30, 1990).

However, Chancellor Donna E. Shalala and University of Wisconsin System President Kenneth Shaw, both of whom also supported an end to the ban, decided instead that the institution would be best served by lobbying the state's Congressional delegation to change the policy (Clark, 1990b; Gross, 1990). They stated in their recommendation to the Board of Regents, the only body with the authority to terminate the university's relationship with ROTC, "that we continue the ROTC program at UW-Madison. At the same time, we join our university colleagues in expressing strong opposition to the current U.S. military policy with respect to discrimination" (Card and Elder, 1989; J.S. Riley, personal communication, August 30, 1990).

In spite of strictly enforced campus and state policies prohibiting discrimination, the University of Wisconsin-Madison made an exception to ROTC, allowing it to remain on campus, over the objections of its faculty (Clark, 1990b; Gross, 1990; Kosova, 1990). When the faculty asked Chancellor Shalala to place a disclaimer in university publications which included information stating that university ROTC programs were in violation of institutional policy and state law, both the Chancellor and System President declined (J.S. Riley, personal communication, August 30, 1990).

On May 9, 1990, the University Council at the University of Pennsylvania adopted a resolution stipulating the suspension of the ROTC program after June 1993, if the DoD were to continue to exclude homosexuals (R.C. Miner, personal communication, August 30, 1990). On May 16, 1990, the faculty at MIT passed a resolution notifying the institutions administration that it was requesting the suspension of the ROTC program after the fall of 1993 if the DoD policy were still in effect (B.R. Maca, personal communication, September 4, 1990). At Syracuse University, the Faculty Military Affairs Committee developed a resolution which was passed on April 30, 1990 by the Faculty Senate, calling for the DoD to change its policy on homosexuals or face the elimination of the ROTC program by June 30, 1994 (R.E. Little, personal communication, August 30, 1990).

ROTC Commanders have asserted that exclusion on the basis of sexual preference is legal and permissible under the authority of federal law which precedes state, local and institutional law, and on the basis that courts have ruled, until recently, that exclusion based on sexual preference was not unconstitutional (Card and Elder, 1989). Administrators assert, however, that at times in the history of the United States, slavery, the return of fugitive slaves, school segregation and the internment of Japanese-American citizens on the basis of their ancestry during World War II were all supported by federal law, only to be later changed (Card and Elder, 1989). Administrators have asserted that qualified homosexual students are placed at a disadvantage while unable to benefit from those same scholarships (Card and Elder, 1989).

William B. Rubenstein, Director of the Gay and Lesbian Rights Project of the American Civil Liberties Union, stated that pressure by institutions on the DoD has been significant because it comes from "the smartest people in our society" (Blumenstyk, 1991). He has cited pressure to remove the ROTC on campuses around the country as the single most important and powerful factor which has pressured the DoD to change its policy (Rubenstein, 1992).

The option of removal has been complicated by the legal responsibilities of land grant institutions, such as the University of Wisconsin-Madison, to offer military instruction under the provisions of the Morrill Act of 1862 (Card and Elder, 1990; Kosova, 1990). However, institutions could remove ROTC presence while still satisfying their military training obligations (Card and Elder, 1990). The Morrill Act states under Section 4 that the funding derived from the sale of lands shall be used for:

> ...maintenance of at least one college where the leading object shall be, without excluding other scientific and classical studies, and including military tactics, to teach such branches of learning as are related to agriculture and the mechanic arts, in such manner as the legislatures of the States may respectively prescribe, in order to promote the liberal and practical education of the industrial classes in the several pursuits and professions in life.

From the development of the Morrill Act in 1862 until the creation of the ROTC in 1916 under the National Defense Act, land grant institutions have fulfilled their obligation through offering compulsory drill to male students and through the development of Departments of Military Science (Hedemann, 1985; Card and Elder, 1989). Under the provisions of the Morrill Act, land grant institutions have the option to remove ROTC from the institution provided they are able to successfully satisfy their respective state legislatures with regard to offering military instruction (Card and Elder, 1989). Therefore, ROTC is not the only way for land grant institutions to fulfill this obligation. Administrators were compelled to take into consideration that public support for homosexuals was not strong, and the impact this would have on politically minded state legislators to support removal.

Further consequences made this option the least realistic. Administrators had to consider that banning ROTC programs from universities would only prevent some students from being able to attend college without ROTC tuition benefits, resulting in a loss of students at a time when funding is badly needed (Collison, 1989). One consequence of banning ROTC units from campuses will be the prevention of capable men and women from attending college and serving their country (W.J. Diol, personal communication, August 27, 1990).

Another consequence of banning ROTC would be the transfer of the primary education for the nation's military leadership to military academies, such as West Point and Annapolis, and to traditionally conservative southern institutions such as Virginia Military Institute and The Citadel, eliminating the ability of more liberal institutions to influence the development of the American officer corps (W.J. Diol, personal communication, August 27, 1990). At service academies, potential officers receive a curriculum much narrower than that found at universities, and at conservative southern institutions cadets would not have been exposed to the liberal environment and wide variety of ideas found at more progressive northern institutions (W.J. Diol, personal communication, August 27, 1990). In addition, if ROTC were removed from the institution, there would have been little, if any, opportunity to pursue a dialogue with the government to end discrimination.

A perspective expressed by one ROTC commander considered there to be little chance of removal, dismissing it as "an old goal espoused by the 60's generation professors who still dislike ROTC. This is a low risk possibility" (M.S. Geoghagan, personal communication, August 24, 1990). In addition, some ROTC commanders did not take academic opposition to ROTC seriously; one Army ROTC commander at the University of Wisconsin-Madison described the institution and city as "an island surrounded by a sea of reality" (J.S. Riley, personal communication, August 30, 1990).

So far, only Pitzer College, one of the Claremont Colleges in California, voted to ask ROTC to leave campus because it conflicted with the institution's non-discrimination policy (Clark, 1990b; Collinson, 1990b). The college faculty voted to remove ROTC from the campus and the president upheld the vote. The policy became effective in September, 1991 (Michaud, 1991).

Distancing the Institution from the ROTC

Administrators opposed to the exclusion of homosexuals as well as the removal of ROTC from higher education have suggested instead distancing institutions from ROTC (Davis, 1990). This point of view manifested in two forms, the physical removal of ROTC from the institution and the disassociation of the institution through policy (Clark; 1990; Davis, 1990; B.R. Maca, personal communication, September 4, 1990).

The concept of physically distancing the institution from ROTC originated at Harvard University, which successfully offered ROTC training at facilities and locations off campus since 1972, when it was relocated in protest of the Vietnam War (Davis, 1990). Since that time, coursework and training have taken place in National Guard and Reserve Centers (Davis, 1990).

The physical relocation of ROTC training satisfied those opposed to the complete removal of ROTC from the institution (Davis, 1990). This response had the added prospect of benefiting both the institution and ROTC cadets. Even while operated off campus, institutions continued to benefit from ROTC scholarships as well as positive relations with the DoD, necessary for funding at research universities (Davis, 1990). At the same time, the interaction with reservists and other active duty officers benefited cadets (Davis, 1990). Physical relocation further eased academic conflict through eliminating the necessity of military officers to hold academic rank, a common source of tension among faculty members (Davis, 1990).

Although travel to such facilities was at times difficult and costly for the cadets and institutions, travel to other sites for ROTC training and coursework continues to be commonplace at small institutions which have historically not produced a large enough number of ROTC graduates to warrant a program of their own (Davis, 1990). From a financial perspective, this policy was appealing to the DoD as well. By reducing the number of battalions while not closing programs at institutions, which would deny opportunity to students interested in ROTC, the DoD was able to commission the same number of cadets upon graduation, at a lower cost (Davis, 1990). Furthermore, centrally located battalions brought together students from a variety of institutions, making ROTC a more appealing option for institutions without a strong ROTC history or for those considering developing an ROTC program of their own (Davis, 1990).

However, this did not address the central issue of ROTC policy, the exclusion of homosexuals (Alderson, 1990; Solomon, 1990). Physical relocation represented only a short-term solution as opposed to developing strategies to change policy in the same manner that the history of discrimination against women and African-Americans was addressed (Solomon, 1990; Williams, 1990). Moving ROTC off campus did not completely resolve that issue.

Administrators who advocated stronger responses argued that only changing ROTC policy would resolve the problem of exclusion and ultimately make ROTC better equipped to produce representative, well educated officers (Solomon, 1990). As representatives of enlightened institutions, administrators were been encouraged to hold their moral ground and confront the issue rather than avoid it (Solomon, 1990).

Rutgers University addressed the conflict between ROTC and institutional non-discrimination policies by distancing itself from the ROTC program it hosts. In November 1989, T. Alexander Pond, Acting President of Rutgers University, ordered the implementation of a report issued by the Select Committee on Gay and Lesbian Concerns, organized by Mr. Pond in February 1988, in response to complaints of discrimination made by homosexual students, directed toward the ROTC program (B.R. Maca, personal communication, September 4, 1990). Under the area of Curriculum and Academic Affairs, the committee made the following recommendations:

> The university must ensure that the joint cooperative academic programs offered by Rutgers in cooperation with outside agencies are in full compliance with Rutgers non-discrimination policy. Programs which do not comply or cannot be brought into compliance should be terminated. Examples include the Army ROTC and Air Force ROTC (B.R. Maca, personal communication, September 4, 1990).

In response to the recommendations of the committee, in August 1990, Rutgers University ceased awarding four merit scholarships to four Army and Air Force ROTC cadets attending the university (Clark, 1990b; B.R. Maca, personal communication, September 4, 1990). Merit scholarships were benefits which paid $1,700 per year in the form of room and board, above and beyond ROTC support (B.R. Maca, personal communication, September 4, 1990). Mr. Pond announced the policy in an open letter to the university support (B.R. Maca, personal communication, September 4, 1990). This was the first time an institution cut ROTC scholarships because of ROTC policy (Clark, 1990b). ROTC students who were receiving scholarship aid at the time of the announcement were not affected.

In explaining this move, William David Burns, Vice-President at Rutgers, stated that the ROTC policy was as "outmoded as their former policy on women and blacks" (Clark, 1990b; Nelson, 1990). He also said that their objective in this move was "not to deny the program to a certain group of students, but to open it to all" (Clark, 1990b). In addition, Rutgers placed in all university literature describing ROTC, a statement that ROTC discriminates against homosexuals and that this contradicts university policy (B.R. Maca, personal communication, September 4, 1990). Rutgers policy prohibited discrimination against individuals based on their sexual orientation (Clark, 1990b). In addition, Mr. Pond joined with presidents of other AAU institutions to write to Secretary Cheney to encourage a change in DoD policy (B.R. Maca, personal communication, September 4, 1990).

Princeton University approached this issue in a similar manner. Among the recommendations a 1990 Princeton University committee made, were revising publications in order to clearly verbalize that the ROTC policy of excluding homosexuals was in violation of the university non-discrimination policy and that the university was opposed to this form of discrimination (Davis, 1990).

Administrators further distanced the institution by lobbying the DoD to increase its efforts to recruit students from graduating classes, minimizing the need for ROTC (Davis, 1990). Such activity would have worked to satisfy the military need for college and university educated men and women, while reducing the need for a military presence on campus (Davis, 1990). Administrators also lobbied the DoD to make more extensive use of summer training programs. Two summer programs instead of one would have reduced the need for military classes at the university, or even have made such classes unnecessary (Davis, 1990).

Many ROTC officers objected to the idea of the institution distancing itself from ROTC (Hatheway, 1990). Physical relocation was considered a low level threat, however, removal of academic credit and budget support was a significant concern, considered to have been a greater threat with a stronger likelihood of occurrence (M.S. Geoghagan, personal communication, August 24, 1990).

A disadvantage to physically moving ROTC off campus was that ROTC cadets would have been less likely to develop fully into free thinking, critically minded officers who are best developed in the higher education environment (Cinquino, 1990). The very presence of military officers in the higher education environment has stimulated the free discussion of the values of the military and higher education in the preservation of national defense (Cinquino, 1990).

> If ROTC enrollments were to drop, the Army would be forced to lower its commissioning standards in order to meet its needs. We have witnessed the results of lowered standards, most recently during the expansion of the officer corps during the Vietnam War. I would suggest that if the officers in common at My Lai had been commissioned through ROTC, we would never have heard of that unfortunate hamlet. ROTC officers, liberally educated, and therefore aware of their larger responsibilities to society and to humankind, would never have allowed a massacre to take place (Cinquino, 1990).

Changing the University Environment

Administrators also approached the problem of ROTC exclusion of homosexuals as part of a larger societal problem with the issue of homosexual rights, and attempted to address the issue through education. In 1990, the University of Minnesota offered its first course on gay studies (McNaron, 1991). English 101 classes in some universities developed writing assignments on homosexuality topics and homosexual writers.

With curricular changes such as these at major universities, it was difficult to ignore the challenges with respect to homosexuals and ROTC (McNaron, 1991). Starting in 1989, Harvard University hosted an annual national conference oriented toward homosexual research and inquiry (McNaron, 1991). That same year, Harvard University appointed an Assistant Dean of Student Affairs to deal with homosexual issues on campus. At Indiana University in 1990, the Dean of Students and Vice-Chancellor began developing an educational task force on gays, lesbians and bisexuals (J.G. Kryway, personal communication, August 28, 1990). In addition, Yale University, Rutgers University, the City University of New York and Berkeley University established homosexual studies programs, while institutions were in varied processes of developing programs of study (McNaron, 1991).

Clinton Compromise

In the spring of 1993, homosexual lobbying groups began pressuring President Bill Clinton to fulfill his campaign promise to improve homosexual rights. However, members of Congress revolted at the idea of allowing homosexuals in the armed forces. The six members of the Joint Chiefs of Staff fueled their anger, lead by its Chairman, General Colin Powell, who strongly resisted the possibility of homosexuals being allowed in the armed forces.

After two weeks of negotiation, the President, with Senate Majority Leader George Mitchell, and Senator Sam Nunn, Chairman of the House Armed Services Committee, agreed to a six month period of policy review, to be accompanied by a moratorium on discharges and disenrollments for homosexuality (Clinton, 1993). On February 3, 1993, the DoD issued revised instructions to the

ROTC regarding enrollment and commissioning of cadets (G. Bond, personal communication, February 18, 1993). These orders included instructions regarding the disposition of homosexual cadets. Under the "don't ask, don't tell" guidelines, cadets were not to be questioned about their sexual preference (G. Bond, personal communication, February 18, 1993). However, any cadet who had stated he or she were homosexual would not be offered a commission (G. Bond, personal communication, February 18, 1993).

Summary and Discussion

The presence of the ROTC on AAU campuses around the country benefited both higher education and the armed forces (Patterson, 1985). Higher education benefited from the scholarship students and funding the DoD has provided. The military benefited from a corps of well educated officers representative of a broad base of the population and more easily able to secure the trust and confidence of a free nation (Malpass, 1985). While these were valuable and beneficial to higher education and the nation, they also were the source of conflict and strained relations with the military establishment.

Institutions responded in a variety of ways to ROTC policy of excluding homosexuals. Administrators who responded supportively cited the many benefits of ROTC to higher education and the nation (Shelton, 1985; Card and Elder, 1989; Davis, 1990; Kosova, 1990). Others tried to remain neutral in order not to interfere with positive and profitable governmental relations as well as to avoid taking a moral position a political issue (Malpass, 1985; Trow, 1987; Davis, 1990; Kosova, 1990; Jaschik, 1993). The banning of military recruiters became an issue in 1982, when law schools at the Universities of California-Los Angeles, Columbia, Harvard, New York and Yale, banned recruiters from their law schools (Norris, 1982). This rose again in 1991 when the Office of Lesbian and Gay Concerns, part of the State Division of Human Rights in New York, ordered the law school at the State University at Buffalo to bar military recruiters from its campus (Blumenstyk, 1991). The complete removal of ROTC from the institution, and severance of all relations with ROTC, was another position institutions considered. However, only Pitzer College, one of the Claremont Colleges in California, voted to ask ROTC to leave campus because it conflicted with the institution's non-discrimination policy (Clark, 1990b; Collinson, 1990b). Administrators opposed to the exclusion of homosexuals in ROTC as well as the removal of ROTC from higher education suggested instead, distancing institutions from ROTC (Davis, 1990).

With the exception of supportive responses to ROTC policy, all other responses shared one common element, changing individual attitudes with respect to how society regards homosexuals. Whether intentional or not, this seeks to indirectly change policy through changing attitudes and values. This is done by sending the community a message that the denial of opportunity to a group of people is unacceptable.

The dispute over the place of homosexuals in the military was more than just a matter of different policy perspectives, it was the continuation of a national battle on homosexual rights (Adair and Myers, 1993). Challenging the military policy regarding homosexuals was one of the central battlegrounds in fighting homophobia at large. Whether one agrees with the policy or not, it was policies such as this which contributed to the emotional and physical abuse and discrimination of homosexuals by sending out an unconscious message that it was acceptable to place these people at a disadvantage and deny them equal treatment (Collison, 1989).

No one knows how many ROTC cadets were disenrolled for their sexual preference (Kosova, 1990). Pentagon statistics are lumped together with those disenrolled for reasons such as failure to meet weight standards (Kosova, 1990). In 1989, 20,178 Air Force ROTC cadets were enrolled across the country (Clark, 1990b; Gross, 1990). Of these, 1,155 were disenrolled for various reasons (Kosova, 1990). The average cost to educate and train a cadet in 1990 was $39,598 (Kosova, 1990). If one were to assume the liberal figure that 10% of these cadets were disenrolled based on their sexual preference, the cost to the United States taxpayers would have been $4.5 million. If similar formulas were applied to all three branches of the ROTC, the cost to taxpayers rises to $9 million at a low point to $18 million at the very highest estimate (Kosova, 1990). This does not even mention the loss of talented leadership that cannot possibly carry a price tag.

Policy makers can learn a great deal from this issue. The level of controversy that developed over ROTC policy makes it clear that administrators cannot underestimate the sense of injustice people will experience if they perceive a group being the victim of discrimination. If that feeling of injustice is strong enough, as it was in this case, it will spill out of the institution into the larger social scene and become a political issue, as was the case in the 1992 national elections.

This is also an issue of institutional priorities, the authority of internal institutional non-discrimination policy over a discriminatory program operating in the institution. Evidence demonstrated that ROTC programs were institutional activities and subject to the authority of the institution (University of Pittsburgh Report, 1991). This was demonstrated by universities providing office space for ROTC staff, salaries, credits, faculty status for ROTC instructors and offering ROTC courses to students who were not cadets (Swartz, 1990; University of Pittsburgh Report, 1991). Allowing outside authorities to determine, in part, institutional policy was nothing new to higher education. This was and is still the case with grant or research funding guidelines, and accrediting agencies. However, by allowing the DoD to operate a program within the institution which openly discriminated against a protected minority, institutions effectively allowed the military to determine institutional policy over the objections of educational administrators, violating institutional policies, posing a threat to academic autonomy. Only by seeking a change in DoD policy, could institutions legitimately and rightfully achieve enforcement of its own policies in its own borders (Card and Elder, 1989).

A disturbing aspect of this policy is the deception and hypocrisy involved. This deception is embodied in examples of senior level administrators publicly criticizing ROTC policy, then privately informing their unit commanders of their support for ROTC. In another instance a university president refused to acknowledge the conflict between an institutional non- discrimination statement, and a policy which openly violated that statement. This type of behavior only served to blanket this issue in more controversy and arouse emotions, making it more difficult to diffuse the problem rationally.

References

Adair, R. and Myers, J. (1993). Admission of gays to the military: A singularly intolerant act. *Parameters*, Spring, 10-19.

Berube, A. (1990). *Coming out under fire: The history of gay men and women in World War II*. New York, NY: The Free Press.

Blumenstyk, G. (1991, January 16). Marine Corps fights attempts to block campus recruiting. *Chronicle of Higher Education*, Vol. 37, 23.

Card, C. and Elder, J. (1991). Position papers addressing eleven frequently heard arguments for allowing ROTC to remain on the UW-Madison campus despite discrimination on the grounds of sexual identity Madison, WI: University of Wisconsin-Madison, Faculty Against Discrimination in University Programs.

Cinquino, J. (1990, June 27). Repercussions of moving ROTC off campus. *Chronicle of Higher Education*, Vol. 36, 4.

Clark, K. (1990, August 29). CSU system faculty joins the anti-ROTC push. *TWN*, 1.

Clark, K. (1990, August 29). Rutgers abolished ROTC scholarships. *TWN*, 10.

Clinton, W. (1993, January 30) "Transcript of President Clinton's news conference." *The Washington Post*, Vol. 74, 12.

Collison, M. (1989, November 8). Many ROTC units under fire because of services' anti-homosexual policies. *Chronicle of Higher Education*, Vol. 36, 37-40.

Collison, M. (1990, March 14). Homosexual ROTC student may be asked to return Army scholarship. *Chronicle of Higher Education*, Vol. 36, 2.

Collison, M. (1990, May 23). College officials urge Cheney to reverse ban on homosexuals. *Chronicle of Higher Education*, Vol. 36, 2.

Collison, M. (1991, February 13). Pentagon to cut number of ROTC scholarships by 20% over five years, starting in 1991-92. *Chronicle of Higher Education*, Vol. 35, 31.

Cushman, J. (1992, November 14). Top military officers object to lifting homosexual ban. *The New York Times*, Vol. 142, 9.

Davis, J. (1990, May 30). Controversies over ROTC could be eased by offering the training at off-campus sites. *Chronicle of Higher Education*, Vol. 36, 3.

Davis, J. (1990). Military policy toward homosexuals: Scientific, historic and legal perspectives. Unpublished masters thesis, Judge Advocate General's School of the United States Army.

Dynes, W., Ed. (1987). *Homosexuality: A research guide.* New York, NY: Garland Publishing.

Fields, C. (1982, August, 24). Six universities could lose defense contracts over law schools' bans on Army recruiters. *Chronicle of Higher Education*, Vol. 24, 6.

Fields, C. (1984, June 6). Law schools told they could lose funds from Pentagon if they bar recruiters. *Chronicle of Higher Education*, Vol. 28, 13-14.

Fields, C. (1984, June 13). Pentagon rules allow it to deny funds to college units that bar recruiters. *Chronicle of Higher Education*, Vol. 28, 15-16.

Gogola, T. (1991, December 2). Ecce Cuomo. *The Nation*, Vol. 253, 693.

Gross, J. (1990, May 6). ROTC under siege for ousting. *The New York Times*, 12-13.

Hedemann, E. (1984). The evolution of ROTC. ROTC/JROTC Clearinghouse, War Registers League.

Jaschik, S. (1993, February 3). Colleges say ending military ban on gays would ease threat to ROTC. *Chronicle of Higher Education*, Vol. 39, 15.

Kosova, W. (1990, February 19). ROTC ya later. *New Republic*, Vol. 202, 24.

Loesler, L. (1945). The sexual psychopath in the military service: A study of 270 cases. *American Journal of Psychiatry*, Vol. 102, 92-101.

Malpass, L. (1985). The benefits of ROTC on campus: A president's perspective. *Educational Record*, Vol. 66 (Winter), 15- 18.

McNaron, T. (1991). Making life more livable for gays and lesbians on campus. *Educational Record*, Vol. 72 (1), 19-22.

Mercer, J. (1992, April 29). New York allows military recruiting as SUNY, despite armed forces refusal to recruit gays. *Chronicle of Higher Education*, Vol. 38, 25.

Norris, W. (1990, March 16). U.S. Army demands gay cadet repays grant. *Times Educational Supplement*, Vol. 906 (11), 4.

Patterson, D. (1985). The University of Southern California: Strengthening academic support systems for ROTC programs. *Educational Record*, Vol. 66 (1), 50-53.

Rubenstein, W. (1991, April 15). The ROTC on campus. Lecture presented at the University of Pittsburgh, School of Law, Pittsburgh, PA.

Shelton, J. (1985). Why military campus programs need campus support in the '80s. *Educational Record*, Vol. 66 (1), 23-24.

Smith, D. (1985). To protect a free society: Maintaining excellence in the military. *Educational Record*, Vol. 66 (1), 11- 13.

Snyder, W. and Nyberg, K. (1980). Gays and the military: An emerging policy issue. *Journal of Political and Military Sociology*, Vol. 8, 71-84.

Solomon, H. (1990, June 27). Repercussions of moving ROTC off campus. *Chronicle of Higher Education*, Vol. 36, 4.

Swartz, S. (1990, August 24). Kansas University meeting to study refusal to ban ROTC commissioning. The Topeka Capital Journal, 1.

Trow, M. (1987, September 30). College trustees should not make moral pronouncements on such political matters as divestment and South Africa. *Chronicle of Higher Education*, Vol. 34, 1-2.

University of Pittsburgh Non-Discrimination Policy Statement. (1990). Pittsburgh, PA.: University of Pittsburgh, Office of Affirmative Action.

University of Pittsburgh ROTC report. (1990). Pittsburgh, PA: University of Pittsburgh.

Williams, C. (1971). *Homosexuals and the military: A less than honorable discharge*. New York, NY: Harper and Row.

Personal Communication

Bond, G., Jr. Memorandum prepared by the Chief of Operations and Marketing Division, United States Army ROTC Command, Headquarters, United States Army, First Region, Fort Bragg, North Carolina, February 18, 1993.

Brown, D. Memorandum prepared by the ROTC Commander at the University of Illinois-Champaign, August 30, 1990.

Diol, W. Memorandum prepared by ROTC Commander at Davison College, Davison, North Carolina. August 27, 1990.

Geoghagan, M. Memorandum prepared by the ROTC Commander at the University of Pittsburgh, August 24, 1990.

Jalette, S. Memorandum prepared by ROTC Commander at The Johns Hopkins University, August 30, 1990.

Kryway, J. Memorandum prepared by the ROTC Commander at Indiana University, August 28, 1990.

Kuhn, L. Memorandum prepared by the ROTC Commander at the University of Kansas, August 28, 1990.

Little, R. Memorandum prepared by the ROTC Commander at Syracuse University, August 30, 1990.

Maca, B. Memorandum prepared by the ROTC Commander at Rutgers University, September 4, 1990.

Minor, C. Memorandum prepared by the ROTC Commander at the University of Pennsylvania, August 30, 1990.

Moore, R. Memorandum prepared by the Presiding Officer of the University Senate at the University of Kansas, May 8, 1990.

Moore, R. and Scott, W. Memorandum prepared by the Presiding Officer of the University Senate at the University of Kansas, May 25, 1990.

Petrick, J. Memorandum prepared by the ROTC Commander at the University of Minnesota, August 29, 1990.

Pullman, C. Memorandum prepared by the ROTC Commander at Washington State University, August 30, 1990.

Riley, J. Memorandum prepared by the ROTC Commander at the University of Wisconsin-Madison, August 30, 1990.

Roberts, P. Memorandum prepared by the ROTC Commander at the University of Iowa, August 29, 1990.

About the Author

Lee S. Duemer
Email: lduemer@averett.averett.edu

Lee S. Duemer received his M.A. (Higher Education, 1993) and Ph.D. (Social, Historical and

Philosophical Foundations of Education, 1996) at the University of Pittsburgh, School of Education. He is currently Foundation and Business Relations Officer at Averett College in Danville, Virginia. His research interest is the history and politics of policy issues in higher education.

Copyright 1997 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://olam.ed.asu.edu/epaa

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey@olam.ed.asu.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Marshall University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Rocky Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
Arizona State University

Dennis Sayers
University of California at Davis

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina--Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher Education

Mary P. McKeown
Arizona Board of Regents

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois--UC

Robert T. Stout
Arizona State University

archives | abstracts | editors | board | submit | comment | subscribe | search

EPAA
• enter the archives • browse the abstracts • the editors • the edit board
• submit article • submit commentary • search • subscribe
volume: • 1 • 2 • 3 • 4 • 5 • 6 • 7

This article has been retrieved 3263 times since April 8, 1997.

Education Policy Analysis Archives

Volume 5 Number 10 April 8, 1997 ISSN 1068-2341

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass, Glass@ASU.EDU. College of Education, Arizona State University,Tempe AZ 85287-2411

Copyright 1993, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

Cultural Differences and the Construction of Meaning: Implications for the Leadership and Organizational Context of Schools

Robert A. Peña
Arizona State University

Abstract The relationships between student achievement, student culture and practitioners' attitudes and expectations were investigated. Student achievement was defined as academic performance but also included perceptions, rationales and explanations for student behaviors and conduct. Student culture described student's Mexican American origins, customs and beliefs. Practitioners' attitudes described how middle school personnel perceived Mexican American high and underachieving students generally, and practitioners' expectations described how personnel interacted and behaved toward Mexican American students. Results indicated that Mexican American students perceived themselves and school personnel perceived these students as different from Anglo students. Mexican American cultural traditions were also perceived as inferior and disadvantageous by high achieving Mexican American students and by personnel. Underachieving Mexican American students generally valued their cultural traditions more positively than high achieving students becoming resistant to learning when these traditions were marginalized in school. Student achievement was also related to student compliance, student appearance, styles in written and verbal communication and practitioners' perceptions about the willingness of Mexican American students to practice and support Anglo norms. These findings are congruent with theories that discuss relationships between student achievement, student culture and practitioners' attitudes and expectations. Theories about school failure occurring less frequently in minority groups that are positively oriented toward their own and the dominant culture were contradicted and not supported in this research.

Introduction

Mirel (1993) notes that during the early 20th century, urban schools were the "jewel in the crown" of the American public school system. Today, unlike their counterparts of almost 100 years ago, Mirel adds that urban schools epitomize the "pessimism and despair" (Edson, 1994, p. 34) of urban decay to the degree that some suggest that they are "not even worth saving" (p. 34).

High dropout rates and academic underachievement are particularly high among urban school students from minority groups according to Mirel. Student alienation due to discrepancies between school cultures and the attitudes and values found in students' homes (Banks, 1993; Brookover et. al., 1982; Edmunds and Fredereksen, 1979; Karweit and Madden, 1989; Weber, 1947) are sited as causes of student under performance. Coleman et. al, (1966) found for example, that academic success and the completion of schooling were due to the "supportive nature" of community life in homes and outside the school, and when a student's values and community relationships mirrored the values and social relationships within the school context. These findings seem deterministic, describing a hegemonic relationship between home and school cultures that frames schools as sacrosanct and student and family characteristics as conducive or not conducive to academic success in school.

In contrast, Peña (1994), Hewlett (1991) and Turnbull and Turnbull (1990) found that school structures that marginalized minorities also led to depressed outcomes for these students. Exclusionary curriculum, scheduling, disciplinary and instructional practices constrained student achievement, limited parental involvement and stimulated antagonistic student behaviors in schools according to these researchers. Scheurich and Imber (1991) also hypothesized that lacking broad community input, policies and practices implemented to benefit underachieving students may also have contributed to their attrition, alienation and underachievement in school.

Empirical and qualitative researchers both suggest that school structures can be deliberately created, maintained, and strengthened through specific approaches to leadership, management, and the manipulation of organizational factors (Bryck, Lee and Smithy, 1990; Newmann,1989; Rosenholtz, 1989). There are, however, three important issues that require further research. The first calls for interviewing minority youth to learn about their self concept and its relationship to achievement in school. The second requires examining the home and school experiences of students in tandem to understand their beliefs about education and in particular, their feelings about social institutions like schools. This approach makes youth and communities rather than schools the primary unit of analysis. The third issue involves analyzing practitioners' behaviors and beliefs to understand how their expectations work with school structures to support and constrain the educational chances, cultures and traditions of minority urban school students.

Sociologists and anthropologists from Emile Durkheim (1984) in his treatise titled *The Division of Labor in Society*, to John Ogbu (1987) in his research on voluntary and involuntary minority groups have found that comparing external social experiences and school organizational characteristics yields information on the values and beliefs of specific groups and how these relate to institutional behaviors and expectations. Analyzing this information may also specify more precisely what organizational features relate most powerfully to the cultural attributes of minority students and to their enhanced achievement in school.

Review of the Literature

Educational theorists attempting to explain minority success and failure in school during the 1980s and 1990s point to what Deyhle (1995) calls "cultural difference" and "sociostructural theories." Deyhle labels James Cummins a cultural difference theorist for example, because of his work and body of ideas on empowering minority students. Cummins (1986) suggests that minority failure and failures in school reform have not significantly altered the relationships between educators and minority youths and between schools and minority communities in his writings. Cummins "central tenet" is that "students from dominated societal groups are empowered or disabled as a direct result of their interactions with educators in schools" (Cummins, 1986, p. 21). His recommendations are that educators change their relationships with minorities to promote empowerment of students which in turn, can lead to success in school.

John Ogbu is described as a sociostructural theorist by Deyhle (1995) because of his writings on economic and political structures, and the academic under performance and dropping out of voluntary and involuntary minorities. Voluntary minorities are described as immigrants "who are doing better in school" and "who have moved more or less voluntarily to the United States because they believed that this move would lead to more economic well-being, better overall opportunities, or greater political freedom" by Ogbu (1989, p. 187). Involuntary minorities describe nonimmigrants who initially were brought to the United States through "slavery, conquest or colonization" (p. 187).

The reasons Ogbu gives for the success and failure of voluntary and involuntary minorities are that immigrants possess a positive dual frame of reference that they use to interpret the "economic, political, and social barriers against them as more or less temporary problems, as problems they will overcome." Involuntary minorities interpret the same obstacles differently and without this frame of reference. Ogbu (1987) suggests that because "they do not have a *homeland* situation to compare with the situation in the United States, they do not interpret their menial jobs as better" or "temporary" (p. 188). For involuntary minorities discrimination is permanent and institutionalized forcing them to look outside of schools and individual effort to collective effort for overcoming barriers to getting ahead. Deyhle (1995) labels Ogbu a "sociostructural" theorist because he argues the reasons for minority student failure lie in the racial, social and economic stratification found in the United States.

In her recent longitudinal study of Navajo students and families on Native reservations, Deyhle (1995) writes that she hopes to "represent the specific Navajo experience" (p. 6). She implies that Cummins' and Ogbu's theories are inadequate because neither addresses "racial warfare" in "both the schools and society" (p. 6).

Deyhle also contends that Anglo teachers and Navajo students engage in "racial conflict," and that Navajos "have substantial ethical disagreements with the Anglo values manifested in the schools and greater economy" (p. 6). This racial conflict also stands for what Deyhle sees as a representation of the integrity of the Navajo culture and figures into the discrimination, subordination, exploitation and to the manufacture of deficit explanations that Anglos create to account for Navajo behaviors in majority dominated schools and businesses.

For Deyhle, the school failure of Navajo youth comes as they have little identity as Navajos and because they are not accepted by Anglos. Deyhle also supports Cummins' (1986) belief that "widespread school failure does not occur in minority groups that are positively oriented toward their own and the dominant culture, that do not perceive themselves as inferior to the dominant group, and that are not alienated from their own cultural values" (p. 32). Deyhle (1995) writes that "Navajo youth who are better integrated into their home culture will be more successful students, regardless of the structural barriers they face" (p. 8). She concludes by asserting that "the more Navajo students resist assimilation while simultaneously maintaining their culture, the more successful they are in school" (p. 8).

These three theories on the success and especially the failure of minority students frame understanding the relationship between minority and majority cultures as crucial to building academic success. Each theory also describes community involvement and acquiring an understanding of the student's community as playing pivotal roles in enhancing school reform and student access and achievement in school. Finally, the authors of each theory insist that what goes on inside the schools, including instructional methods and the kind of curriculum taught, are very important for minority student success.

Where each of these theories lingers is in explaining the success and failure of minority students with similar cultural, community and school backgrounds. These theories do not account for students who reside together in the same community, share the same cultural background, have the same teachers and like schedules of classes, experience the same instructional methods and curriculum in the same school, speak a version of English at home, whose home language and culture differ from those of the school and wider society yet who also show high and underachievement in their classes .

Cummins (1986) posits that minority language incorporation, community participation, enhanced intrinsic motivation and the professional acting as an advocate for minorities are four key dimensions that operate on a continuum and promote the empowerment of students on one end while contributing to the "disabling of students" (p. 21) on the other. Ogbu (1987) suggests that variability in minority school performance at the individual level can be traced to differences in cultural models: to the initial terms by which the minorities were incorporated into U.S. society; and by the way minority students interpret their initial incorporation and their subsequent treatment by white Americans. Deyhle points to the importance of reservation life and to the preservation of traditional culture for Navajos as contributing to failure in public schools that stress competitiveness and individuality.

In each discussion, these theorists neatly explain how the dominant culture diverges from and seizes the weaker less traditional culture. Cummins, Ogbu and Deyhle also suggest that superior integration in the school and community is necessary for increasing minority academic achievement and greater success overall. This study consequently tests these theories by attempting to understand why students unable and able to maintain their cultural connections nonetheless contradict and prove successful and unsuccessful respectively in the Anglo world of schooling. This study hopes to expand previous understandings by analyzing the success and underachievement of twenty Mexican American students that live in the same feeder neighborhoods and are enrolled together in a single public middle school in a state located in the Southwest.

Theoretical Framework

Based on the studies described earlier, minority student achievement may be improved by making school factors more relevant to student backgrounds. School attempts to enhance school membership, teacher expectations, educational engagement and school support presumably yield improved student performance and outcomes (Peña, 1995; Wehlage, 1989; Wehlage, 1986). Researchers also agree that school traditions that do not agree with students' cultural attributes will adversely effect membership, instruction and the disciplinary climate in schools (see Erickson, 1987, McNeil, 1986 and Willis, 1977). Consequently, this researcher proposes to examine the school and community experiences of high and under achieving first generation Mexican American working class students to understand how these students define themselves, education and success in schools and in their community. This examination may generate understandings on how attitudes and school cultures support and constrain the achievement and behaviors of these students and members of their ethnic and racial peers. Mexican American pupils describe first generation students who have some English proficiency skills and have taken up permanent residence in the United States.

Although there is significant variability among first generation Mexican American students from working class families, individuals from these groups may nonetheless share "underlying cultural patterns that influence their behaviors and beliefs" (Deyhle & LeCompte, 1994, p. 156). Labeled "cultural boundaries" by Erickson (1987), studying these patterns may give evidence of different "ways of growing up," "raising children," and "evidence of different cultural standards of appropriateness" (Deyhle & LeCompte, 1994, p. 156). Studying the home and school experiences of high and under achieving Mexican American students then, may explain how they define themselves, how they interact with peers and school personnel, and what attitudes and behaviors these students exhibit that enable them to succeed and fail in school and in their communities. Community in this context describes a specific external location where persons live, share daily interactions and a location that is contained by school boundaries and common to the students included in this study.

Methodology and Sources of Data

Data generated through depth interviews, document analyses, and participant observation were analyzed using constant comparison and methods taken from grounded theory (Glaser, 1978; Glaser and Strauss, 1967; Strauss, 1987). Constant comparison describes the simultaneous collecting and analyzing of data for their refinement, categorization and integration into a coherent theory (Taylor and Bogdan, 1984). High and underachieving Mexican American middle school students and school personnel who routinely interacted with these students were interviewed, while guardians and relevant members from the community were observed.

Data collection started upon acquiring the recommendations of administrators and teachers for "ten high achieving and ten underachieving Mexican American students" to interview. Decisions for expanding and including others were based on snowballing techniques where interviewees recommended additional participants, on the development of themes, and on the emergence of data saturation or the point at which information collected became redundant (Bogdan and Biklen, 1992; Glaser and Strauss, 1967). In total, 20 students were nominated (ten high achieving and ten underachieving Mexican American students) and participated to the completion of this study. Additionally, 12 teachers and two middle school administrators were interviewed and observed.

The formation of questionnaires, elements, themes and supplemental data collection instruments for document analyses and observations were guided by the theoretical framework described earlier and by prior social science and anthropological research (see Cummins, 1986; D'Andrade, 1984; Deyhle, 1995; Erickson, 1987a; Erickson, 1987b; Goodenough, 1981; Geertz, 1973 and Ogbu, 1987). Analyses of discussions with different respondents, documents and observation notes were also employed to understand Mexican American student self perceptions, their perceptions of schooling and how school policies, practices and practitioners' perceptions relate and contribute to their success and failure in a single middle school located in the Southwest.

Findings

Analyses of the data indicate that the ten high achieving Mexican American students demonstrated attitudes and behaviors that were distinct from their underachieving peers in and out of school. High achieving students were compliant with demands placed on them by teachers, middle level structures and other requisites for social acceptance and achievement in school. These students also framed meeting school demands as more important and personally satisfying than pursuing ethnic membership. High achieving students also viewed their cultures as embarrassing more often. These students described experiences in Mexico, at home and characteristics of language and culture as impediments to fitting in, gaining social acceptance and their achievement in school.

Underachieving Mexican American students in contrast, were generally less compliant and more resistant to school customs that agitated and marginalized their own cultural traditions. These students placed cultural membership before achievement in school more often, attaching greater import to cultural knowledge and integrity than to being compliant and making friends and grades in school. Social acceptance emerged for these students through relations with family members, close peers and community members with similar values in both informal and middle school settings.

Analyses related to teachers' perceptions indicated that educators spent little time and possessed scant knowledge of their Mexican American students backgrounds. Practitioners also felt that higher achieving students possessed a clearer sense of personal identity than underachievers, and that these students were more willing to adapt to and prevail over different demands that might be perceived as culturally antagonistic by minority students. Teachers also agreed that high achievers demonstrated greater fluency in Spanish and English than underachievers, and greater mastery in transferring and adapting prior experiences and understandings to unfamiliar concepts and traditional instructional methodologies.

Teachers added that underachieving Mexican American students seemed less capable of expressing their thoughts and reasoning about prior experiences in a thorough and orderly way. They felt that underachievers demonstrated sporadic flashes of thought in school while demonstrating a cultural rift, unable to integrate their experiences on Mexican and US soil. Teachers also concluded that this cultural rift prevented underachieving students from applying prior educational experiences and knowledge for making meaning of instruction and expectations in traditional US schools. What follows is detail on higher and underachieving Mexican American students, the strategies they used to make sense in school and in their communities, and school and community factors that supported and constrained their school success. In this context, the nature of school success and failure is considered using two frames (Erickson, 1987b). These frames refer to the ways that students succeed and fail to achieve in school and in their community, and to the ways their school and community support and fail Mexican American middle school students.

Understanding High and Underachieving Mexican American Students

Explanations for success in and out of school were organized under three domains. The first, or personal domain fixes explanations to the students, their families, their backgrounds and to students' lifestyles. The second or interpersonal domain attaches success and failure to students' peer and social relations. The third category labeled formal and informal domains contains fixes student success and failure to power configurations and the interplay between school and community characteristics.

Personal Domain

Students explanations for their high and underachievement in school were based upon assimilationist and cultural resistance ideologies. The high achieving students understood they were different from Anglo teachers and students and that academic achievement required them to "work harder to prove we all aren't dumb and we could do it [achieve] too." These students also perceived they "have to be better than everybody else all the time because you want to be like them when you're in school," because they needed "to fit in," and because "you want your teachers to like you" and "have teachers help you out." One high achieving student noted that "everyday you remember you're not from here even if you are, and then your mother and father talk different and are not from here and that you're really not as good and maybe don't look like you belong in this school." Another student recalled concealing her anger and embarrassment over Mexican American students being singled out and treated unfairly in class:

> "Mrs. Thomas likes to put the Mexican's against the Anglo kids all the time and I really hate when she does that because it's not that right. For recess she treats us like little kids and she makes us go to the door and line up and be quiet. The kids with the green eyes go first, then the kids with the blue eyes then if you got brown eyes you go last sometimes. Then another time in spelling Lucinda got marked down because she didn't spell her word loud for Mrs. Thomas to hear her. Then when the students said they didn't hear Judy talking loud enough either then Mrs. Thomas told everybody to be quiet and then she said to Judy to spell her word over again and louder this time. I said that wasn't fair and Mrs. Thomas looked angry at me and I could feel my face turning

> all red inside you know because everyone was looking at me. Then she said we weren't at home and if we didn't behaving right she was gonna cancel everything for the spelling contest and pick the winner for class by herself."

Underachieving Mexican American students in contrast were less interested in demonstrating compensatory behaviors and making a positive impression on their Anglo teachers. Like the ten high achievers, the ten underachieving students understood they were different and did not measure up to Anglo teachers, students and school norms. These students also felt they could achieve and excel in school, but they were more often unwilling and resistant to provide answers in class when they perceived they were being singled out because they were Mexican, Mexican American and different.

One student recalled being "picked on by the teacher to say who was Jackie Robinson and what was Jackie Robinson famous for." This student correctly explained to the interviewer that "50 years ago he [Jackie Robinson] was the first Black man to play in baseball" but added that he "didn't answer" and "went like this [raised his shoulders] like I didn't know" because he perceived "he [the teacher] asked me because I'm Mexican and we're supposed to know about sports and who was first and shit like that."

Another student recalled when he and his classmates were "pressed on" or "hassled during PE [physical education] because we were hanging and talking in Spanish on the side and laughing and we didn't want to get into it [play basketball] and all dirty and everything." This student explained that he believed "the teacher got mad because he thought we was talking about him" and "we weren't ready for class." When asked to tell what happened next, the student answered that "they [his teachers] forgot about me" and that he "had to sit in the office for making a face at him [the teacher] or some other shit for over a hour."

Finally, a third student said that "everybody knows you have to give up being Mexican to do good in this school." When asked to explain this student added:

> "...it starts right at the beginning of the year when everybody tries to be real nice. They hook you up in the same homeroom with the same teachers because they think you don't know nothin and you're stupid and you don't speak English the right way or something. And they talk real loud and slow so you understand what they're saying just because we're from Mexico. It's like the school already made up their mind about us even before we got here that we're dumb and if we change in school like they tell us then we'll stay out of trouble and we'll make it okay. I guess they want us to act different like our families didn't come from Mexico or something and we should be like we're American in school like that's something right or whatever."

Explanations by school personnel for the success and failure of Mexican American students that were also attributed to the personal characteristics of students related to congruities and incongruities in individual versus formal (school) styles of learning. Teachers believed that high achieving students jockeyed for high grades, praise and recognition in school more often than underachievers for example, because these students attached greater significance to school and personal recognition than to benefits that might accrue from cementing cultural membership for themselves.

Underachievers, according to administrators and teachers also demonstrated loyalty to their cultural traditions and origins more often, becoming upset and resistant to learning in school when cultural characteristics were ignored, did not match and were handled negatively by educators. Finally, teachers also believed that higher achievers were more pleasant, willing to please teachers and demonstrate positive behaviors than their underachieving peers who seemed less trusting and more cynical about how "Mexican" and "Mexican American" traditions were treated in school.

Evidence of compensatory and resistant student behaviors emerged during interviews with teachers and during observations of instruction and observations of classroom patterns of interactions. Teachers explained that it was "very important," "real important" and "more important for high achieving Mexican American students to get [good] grades in school" for "getting into college," "for making some money," "for making lives for themselves," and for these students "to be liked by their teachers." Two teachers added that "high achievers and their guardians concur that it is important to succeed in school in the United States" and "they understand it's real important to make the effort to get along with people."

All teachers were also impressed with the "industry" and "more pleasant demeanors" of high achievers mentioning that these students were "appropriate" in dress and "neat" when completing assignments. These teachers also explained that underachievers were "more demonstrative," "insubordinate," "less neat," "messy" and that their assignments were "not always finished or handed in on time." Higher achievers also completed "work early" on occasion even doing additional work while underachievers behavior and attendance was described as "less reliable" and "not as friendly" by teachers.

Analyses of field notes, specific verbal exchanges and samples of students' writings similarly indicated that teachers praised students for style in the forms of precise language skills and in writing mechanics. Teachers also described their appreciation for students who "knew things," "were always in class" and for students that "did not interrupt" and apparently placed fewer demands on teachers.

In contrast, underachievers were described as students "who constantly needed supervision and guidance" with "poor mechanics in writing." These students were described as "silent," "unmotivated" and "car[ing] less about standard pronunciation." Teachers also felt that underachieving Mexican American students made "less effort to correct errors," "to learn from their mistakes," and that these students were less skilled in "transferring and applying knowledge," "synthesizing information" and "using analytic and upper level thinking skills" than their high achieving peers.

Interpersonal Domain

Students descriptions of their interpersonal relations with teachers, peers and members from their community were similarly influenced by their inclinations toward assimilation and resistance, and their beliefs about the supportive and non supportive characteristics of their Mexican American culture. The high achieving students actively pursued recognition in school for example, choosing to associate with other high achievers and recipients of school accolades regardless if they were Mexican American or not Mexican American students.

These high achievers also seemed more eager for competition for praise, higher test scores and higher averages on first term report cards than for affirming their cultural identities. For them, academic achievement and positive social relations in school became hard earned wages that took on a transactional significance. Each A or B grade and word of praise was like another dollar adding up to a rite of passage for membership in a student association or "college Greek house" with other high achieving students. Their cultural background on the other hand, was a constant impediment; a reminder to these students that they were different and not wholly accepted in the formal order of school.

One high achieving student explained that "we [Mexican students] have to be better all the time to show we're good as Anglos and we belong here." A second student said "I try and be the best in everything I do. In school, in PE [physical education] too." This student explained that "sometimes the kids tease me because of my hair or my skin or something, or another time when my mother spoke Spanish and she came to get me... so I get good marks and everything and that I'm nice and just like they are so I get along better with them."

Other high achievers said that classmates were "nice," "ask[ed] for help," "think you're smart," "walk together" and "pick you for doing things" if they earned high grades and praise. Finally, one high achieving fair complexioned student shared his strategy this way for fitting in with others:

> "When I'm alone and not with anybody I don't tell people that I'm a Mexican right away. My last name is Mexican but a lot of people don't know my name before so I don't say nothing and they think I'm American or Italian sometimes. Then sometimes when my friends in school get on me about my shoes or my clothes or what I bring to lunch or whatever, I pretend like it doesn't bother me and I make fun too. Then sometimes I shift what we're saying and talk about another thing or another classmate or whatever. I never had too many people come over my house because they always say my mother talks too fast so they don't understand what she said."

Underachieving students in contrast, neither pursued recognition for academic performance nor did they seek association with high achieving students. For them, high achievement was like "being Anglo" or Anglocanized with negative consequences for their Mexican identity. Additionally, underachievers more often gravitated rather than actively moving toward peer and social relations in school. Their social circles seemed to include fewer students and to include more trusted peers from their local neighborhoods and community.

Specific data on social patterns for underachieving Mexican American students emerged during interviews and especially during observations of these students in school, their homes and in their surrounding neighborhoods. These students seemed uncomfortable in school more often than high achievers yet more comfortable out in their neighborhood communities. Pregnant with expectation as though they were waiting for someone or something to change their lives, these underachieving students often belonged and fit best in tight knit social circles. For these students, school was a challenge where their personal faith and cultural loyalty was regularly tested while life in their homes and neighborhoods brought predictability and ease. Interactions with Anglos and high achieving Mexican American students were usually guarded and suspicious while their noncompliance in school was also proof of their cultural integrity and loyalty to their Mexican roots.

One student described routinely "go[ing] late [to school] to get out of confrontation with [the mathematics teacher] during first period." This student explained that the mathematics teacher "...gives homework everyday even on the weekends then when you're in school she makes you get it out so she could come to your desk and give you a hard time if you don't have it." This student added that she did not know "why you should have to do the homework all the time if you get it," and that "doing homework" and "being good in class is for the Anglos and the wanna-be's."

A second student described high achieving Mexican American students as "trying to be so white they're squeaky clean." This student explained that high achievers worked for grades and "try to talk English good because they want people in school to like them because they don't like being Chicano." This student added that "people think it's bad because we're dumb and don't have no friends in school but they [high achievers] don't have no friends in the [neighbor]hood." This student went on to explain that "they [high achievers] don't know what's going on" and that "you never see them outside or in church with anybody or with any friends out of school because they turned their back and forgot who they are for real." Finally, this student also warned that "when something happens and they aren't doing good... then we'll see what Anglo friends they got because they won't have any."

Other underachieving Mexican American students echoed these statements, explaining that high achievers were "fools," not "liked," "disrespected," "chumps," "dogs" and "ghosts" in their communities because they "disappeared," were "invisible," did not "come outside ever," were "not respected" and "never did anything in the neighborhood except for go to the store once in a while." These students further explained that they preferred making and having friends in their community because "there's no front," "you could be yourself," "there's more trust," "people [in the neighborhood] know what's going on," "everybody's the same," because these students "like the neighborhood" and because "you could see someone [from the neighborhood] in the eye and know what's goin on with them."

Finally, one underachieving student said that:

> "...it's real hard to be good in school and in the neighborhood at the same time. It seems like it starts real early like when you're in third grade or second. Your mother and your father they're on you all the time to do good in school and to get make better grades than they did, but then you're torn up. You see the way the Anglos treated better in school better and how when you do the same thing but it doesn't make matter. Then you come home and all your mother and father tell you is you have to do this and it's gonna be okay or whatever and then you start to hate it and that you know because it isn't. You go with your friends and your friends come over and they hate what happened in school just like you do too. And then it's all bullshit all over again like you're dirty or something and the good [Mexican American] students are dirty too except they don't know it or something and their clean on the outside and the Anglo's are the only ones that are good. It's like everyday they [teachers] already made up some secret about us and that we're Mexican so we got to remember that everyday wherever we go in school. I remember it because I want to because I'm proud to be Chicano. I don't need nobody to tell me. I want to be proud and my mother and father and sisters they're proud too, but not the teachers... It's like they have some problem or something before they even know who you are and then your mother and father want you to do good too."

Explanations provided by school personnel for students' social patterns were similarly attributed to the compliance and resistance of students and to students' attitudes about their Mexican culture. Teachers generally believed that high achieving students were "more pleasant," "sweeter," "comfortable" and "at peace" with their Mexican culture for instance, than were underachievers who were "less forgiving," "bitter," "angry" and "more combative" when they perceived their cultural traditions were being insulted.

Teachers also described high achievers as "happier" and from "better more supportive homes." These teachers added that high achieving students had "more desirable" and "greater numbers" of "white and Mexican American" friends than underachievers who tended to associate with "other poor performers" and "less friends" who are "usually Mexican" and "friends that are usually in trouble too." Finally, teachers also believed that higher achievers were more likely to "succeed" and "make something" of their lives than were their underachieving peers who "seemed less trusting" and experienced "more trouble making more than their few friends."

Data supporting teachers' accounts of the compliant and resistant nature of students emerged when teachers described the attitudes of their Mexican American students. One teacher commented that "it's easier to enjoy students with a more pleasant attitude than those who behave suspiciously." Another teacher explained that high achieving students "have more friends because they apply themselves more and have more to offer in school." A third teacher added that "high achievers extend themselves and are willing to meet others half way" while a fourth said students "learn at home it's real important to make the effort to get along with people inside and outside of school."

All teachers also agreed that making friends was "more important" for high achieving students. According to these teachers underachieving students more often "drifted" from one friendship to another." These teachers added that underachieving Mexican American students "spoke less to adults and other children," were more often "introspective" and "mysterious," and that these students "have low self concepts," "low confidence" and "immature social skills." Underachievers were also described as "awkward" and "uncomfortable" when being addressed by teachers.

Analyses of field notes compiled largely through observations revealed that teachers touched and responded pleasantly more often to high achieving than to underachieving Mexican American students. Like their students, teachers also seemed more comfortable and at peace with high achievers and more awkward and less forgiving with underachieving students. Analyses of notes indicated that high achievers were left unsupervised more frequently for instance than were underachievers, and that teachers were more hasty and severe when disciplining underachieving students.

Teachers scolded, showed their appreciation and attempted to correct high achieving Mexican American students who they felt behaved inappropriately in class on occasion while choosing to talk loudly, yell, crowd, become physical and remove underachievers for interrupting classroom instruction.

Formal and Informal Domain: Formal and Informal Cultures

The formal and informal domain is also labeled formal and informal culture in this manuscript. Formal cultures describe the customary beliefs, social forms and institutional structures that a particular group of students or individuals encounters in school. Informal cultures describe the same characteristics, groups and individuals but as they intermingle, create meaning and are defined and redefined in an informal setting.

Formal and informal cultures are conceived here not as static but as active as groups and individuals are routinely and significantly affected by environmental contingencies. These cultures may be marked by "underlying cultural patterns" (Deyhle and LeCompte, 1994, p. 156) that characterize group and individual behaviors and beliefs, and by environmental factors that collide and struggle with these patterns and against one another to establish social control and a sense of equilibrium in schools for example.

This notion of formal and informal cultures then is conceptualized as an inchoate number of variables leading to a particular result rather than as a postulated outcome or event. A similar description of culture as process is implied in Harrington's (1962) *The Other America: Poverty in The United States*. In this influential book (Spring, 1976), Harrington introduces the "culture of poverty" explaining that trapped within a "vicious circle" with inadequate nutrition, medical care and lost wages, the poor get sick more often while their sickness stays longer. This image synthesizes the characteristics of people and their lifestyles with environmental factors to establish that when combined, a culture of poverty is made. In short, the individual's personal characteristics and the characteristics of their environment conspire to economically disable them in this case. The individual's personal attributes in isolation are neither adequate to describe nor to confine them then to the culture of poverty.

The notion of formal and informal cultures and the process previously described is hypothesized to be violent and deleterious as nontraditional and weaker cultural orientations hide, adjust, resist or become trampled by stronger more traditional understandings in a formal or an informal setting. Additionally, informal understandings may not prosper and survive in a formal environment and formal knowledge may wither and die on the vine in more informal environs.

Taken together, this struggle for legitimacy, control and social equilibrium becomes a chaotic yet systematic attempt to establish order where threats to that order constantly emerge. This struggle between formal and informal cultures may also be imbued and bereft of morality and the human spirit at the same time, depending on the relationships and organization of groups and individuals, and social, political and economic configurations of power.

This discussion comes from the previous research on high and underachieving Mexican

American students who all together seem required to regulate formal and informal cultural understandings in a formal middle school setting, and who also are all required to weigh and manage these pursuits in their local communities. Further analyses of the data collected indicate for example that none of the 20 students interviewed was comfortable and flourishing in both their school (formal) and community (informal) environments. Based upon their sense of personal efficacy, students would seemingly achieve or resist in one setting, and struggle and flourish in the other. Success in school came more readily for those willing to understate, separate from or deny their Mexican culture. Students who emphasized their Mexican cultures on the other hand, experienced low expectations, failure and hardship in school while experiencing respect and fulfillment in their community more often.

Further analyses of interview and observation data collected also indicate that high achievers generally preferred school experiences to life in their neighborhoods while underachievers preferred the comforts found in the community. For high achievers, school appeared to provide rationality, a routine and to bring certainty to their daily lives. Expectations on thinking, dress, scheduling, behaviors and rewards were clear in school but muddied when high achievers returned to their neighborhoods. Expectations in school for underachievers on the other hand, were too severe requiring them to change their intellectual approaches and to cash in their cultural understandings for a chance at high grades and assimilation. At home in their neighborhoods, underachievers felt they could think and act for themselves, make sense of local activities, events and behaviors, detect and understand the glances of neighbors, and empathize with passers by on the street. Further analyses of data collected in the middle school indicate that teachers usually preferred higher achieving Mexican American students. Teachers often approved of these students more because they were compliant, hard working, reliable and because high achievers interrupted less and placed fewer disciplinary and book keeping demands on them. Teachers also judged high achievers as superior analyzers and evaluators of knowledge, more popular, better socially adjusted and more concerned about achieving a better future without necessarily testing students higher order thinking and without observing students in their neighborhoods. Finally, teachers also described the parents of high achievers as more supportive than the parents of underachieving students without talking to them or visiting their homes.

Discussion

Conclusions drawn from the data collected supports earlier assertions on the importance of understanding the relationship between minority and majority cultures while adding discourse on formal and informal cultures and on the importance of considering the school and home communities of students perceived to be different. Results from this study also gave no evidence that minority group members that are positively oriented toward their own and the dominant cultures are better prepared to resist failure in school. In contrast, students who viewed their Mexican American culture less favorably achieved in school and were less accepted in their communities. Those who emphasized their Mexican American culture underachieved in school and flourished at home. Finally, this study weighed the value of making school processes more culturally relevant finding that the promotion of cultural traditions in school held promise but did not benefit all members of a particular minority group equally.

Analyses of the data collected in this research suggest that it is equally important to understand the relationship between minority and majority cultures, and to understand the interplay of these in both the school and community. This means that educational leaders and school practitioners become knowledgeable of minority cultural traditions, and that these individuals become more reflexive in their thinking about culture. In other words, a fuller understanding of cultural differences may require experiencing them in and out of the formal educational setting, and perhaps experiencing what it means to be different in a predominantly minority context.

Fluency in school policies and being an effective administrator of school procedures that reflect Anglo preferences solely is not conducive to supporting achievement and minority culture, and is akin to asking members of minority groups to support Anglo school structures and traditions they are unaware of and do not fully understand. On the other hand, neither does full immersion in formal and informal settings guarantee that one will become an insider or that changes in personal attitudes and patterns of discrimination will emerge. Conclusions on the relationship between full immersion programs and individual's perceptions of cultural differences requires additional research. Further study of district transportation and zoning policies are also needed to understand how these support and limit knowledge about what is appropriate and inappropriate in the school and community context.

As noted earlier, conclusions about minority students being better prepared to avoid school failure by holding positive orientations of both their own and Anglo cultures were also not supported in this research. Analyses indicated that the academic performance of students was value-laden and largely related to practitioner's judging habits. Grades appeared to be used as a means for rewarding, penalizing and separating students, while achievement was measured according to students' attention to detail, writing and speaking habits, physical appearance, and minority student's attitudes about Mexican American and Anglo cultures.

Teacher habits in assessment also led to untenable conclusions about the intellectual makeup of students and the supportive and non supportive nature of students' backgrounds. Teacher made tests and styles of questioning did not measure students' application, analytic and evaluation thinking skills for example, although underachieving Mexican American students were judged less competent in higher order cognition.

Errors about the readiness of students to benefit from learning and about the willingness of families to help students learn were also made as teachers decided that high achievers and their guardians naturally valued learning more than families with underachieving children. This is not to say that the readiness of the students in this study could not be benefited from compensatory programs. Instead, analyses suggest that because the range of student cognition was not adequately addressed, accurate decisions about effective pedagogy, curricula and school reform also could not be made. This finding means that administrators skills in instructional leadership and supervision need refinement so they can help practitioners become more competent in teaching and assessing students' higher order thinking. This also requires that district supervisors and researchers play a larger role in understanding possible relationships between culture, learning styles and student assessment.

Finally, while the practice of making school policies and procedures more culturally relevant appears to hold promise, analyses conducted for this study contradicted earlier writings by showing that the random promotion of specific cultural traditions in school did not benefit all members of a particular minority group equally. High achieving students generally viewed their cultural traditions as embarrassing and as impediments to their acceptance and achievement in school. Underachievers valued their cultural identity more producing resistance to learning, alienation from other students and conflict with teachers. In contrast, high achievers also enjoyed their home communities less feeling insecure and uncertain more often than underachievers who generally felt less scrutiny, more belonging and more comfortable at home.

Implications for theorists, education leaders and the organization of schools require that they become knowledgeable about the relationship surrounding student self concept, social acceptance, culture and the achievement of minority students in school. Analyses of the data collected indicates that minority students value fitting in with others in one setting or another, and that their self concept, willingness to participate and freedom to learn are constrained to the extent they feel alienated from their peers, their community and their cultural understandings.

In this context, being Mexican American also meant being different in school and that this difference was perceived by students and educators to mean naturally inferior to Anglos. High achievers worked hard to gain school membership by deferring their cultural identities while underachievers worked hard to keep their cultural identities and membership at home. Understanding how to promote self concept, acceptance and belonging in school and in the external community seems important for improving students' academic achievement. This suggests that researchers and practitioners become more compassionate and knowledgeable of the relationship between formal and informal cultures, and the implications of this relationship for helping youths feel better about themselves, achievement and their place in school.

Conclusion

Like other research, this study ends prematurely probably raising more questions than it answers. Early on, it included highly general causal theories by Cummins, Ogbu and Deyhle that link school success and failure to cultural differences, sociostructures, and racial conflict. Then, it explained that these theories were inadequate demonstrating how students that fit these models nonetheless achieve in school and in their home communities. This inquiry consequently expands on the literature reviewed while also serving as a warning against simple explanations to challenging issues. It also asks that researchers think "more self-consciously about the philosophical and political implications and meanings" (Scott, 1988, p. 134) of the theories they endorse.

Next, results coming from this study reminded readers how classifications by culture, ethnicity and race may be based on delusions (Husband, 1982) as they lack scientific validity and are largely informed by socio, political and economic pressures. Students' attitudes on fitting in at school or in their home communities, and teachers' behaviors toward Mexican American students in this research, related to their perceptions of difference.

A positive definition of Mexican American culture rested on the desire and ability of high achievers to think and act "normally," or as the dominant Anglo group in the school believed they should. Negative definitions of Mexican American underachievers emerged because their behavior was perceived as resistant and antagonistic, and because their culture seemed antithetical to the dominant Anglo culture in school. This suggests that the Anglo culture was accorded primacy in school while the Mexican American culture was secondary. This also suggests that the educational experiences of the students included in this study were largely based on cultural contrast and subjugation rather than from some cultural interdependence.

Future research on student achievement and failure must continue with a deconstruction of cultural relations and how difference is constructed in school. Future research must also strive to assess the interdependence of cultures in and out of schools to determine how schools can foster cultural harmony and intellectual, social, political and economic gains for all.

References

Banks, J & Banks, C. (1993). *Multicultural education: Issues & perspectives*. 2nd ed., Needham Heights, MA: Allyn & Bacon.

Bogdan, R.C. & Biklen, S.K. (1992). *Qualitative research for education: An introduction to theory and methods*. 2nd ed., Needham Heights, MA: Allyn & Bacon.

Brookover, R. (1985). Can we make schools effective for minority students? *Journal of Negro Education, 54*, 3: 257-268.

Bryck, J., Lee, V., & Smithy, C. (1990). Are head-start effects sustained: A longitudinal follow-up comparison of disadvantaged children attending head start, *Child Development, 61*, 2: 495-507.

Coleman, J.S., Campbell, E.Q., Hobson, C.J., McPartland, J., Mood, A.M., Weinfeld, F.D., & York, R.L. (1966). *Equality of educational opportunity*. Washington, DC: US Government Printing Office.

Cummins, J. (1986). Empowering minority students: A framework for intervention, *Harvard Educational Review, 56*, 1: 18- 36.

D'Andrade, R.G. (1984). Cultural meaning systems. In R.A. Schweder & R.A. Levine (Eds.) *Culture theory: Essays on mind, self and emotion*. Cambridge: Cambridge University Press.

Deyhle, D. (1995). Navajo youth and anglo racism: Cultural integrity and resistance, *Harvard Educational Review, 65*, 3: 403-444.

Deyhle, D. & Lecompte, M. (1994). Cultural differences in child development: Navajo adolescents in middle schools, *Theory into Practice, 33*, 3: 156-166.

Durkheim, E. (1984). *Durkheim: The division of labor in society*. Introduction by Lewis Coser. Translated by W.D. Halls. NY: The Free Press.

Edmunds, R. & Fredereksen, J. (1979). Effective schools for the urban poor, *Educational Leadership, 46*, 5:29-32.

Edson, C.H. (1994). *Crisis in Urban Education*. Albany, NY: SUNY Press.

Erickson, F. (1987a). Conceptions of school culture: An overview, *Educational Administration Quarterly, 23*, 11-24.

Erickson, F. (1987b). Transformation and school success: The politics and culture of educational achievement, *Anthropology and Education Quarterly, 18*, 335-356.

Geertz, C. (1973). *The interpretations of cultures*. New York: Basic Books.

Glaser, B.G. (1978). *Theoretical sensitivity: Advances in the methodology of grounded theory*. Mill Valley, CA: Sociology Press.

Glaser, B.G., & Strauss, A.L. (1967). *The discovery of grounded theory. Strategies for qualitative research*. Chicago: Aldine.

Goodenough, W.H. (1981). *Culture, language and society*. Menlo Park, CA: Benjamin/Cummings.

Harrington, M. (1962). *The other America: Poverty in the United States*. NY: The Macmillan Company,

Hewlett, S.A. (1991). *When the bough breaks: The cost of neglecting our children*. New York: Basic Books/Harper Collins.

Husband, C. (1982). Introduction: Race, the continuity of a concept, In *Race in Britain: Continuity and Change*, Edited by Charles Husband. London: Hutchinson.

Karweit, N., & Madden, K. (1989). The effects of a story reading program on the vocabulary and story comprehension skills of disadvantaged prekindergarten and kindergarten students, Microfilm: ED313655.

McNeil, L. (1986). *Contradictions of control: School structure and school knowledge.* New York: Routledge & Kegan Paul.

Mirel, J. (1993). *Rise and Fall of an Urban School System: Detroit 1907 - 1981.* Ann Arbor: U. of Michigan Press.

Newmann, F.M. (1989). Organizational factors that affect school sense of efficacy, community, and expectations, *Sociology of Education, 62*, 4: 221-238.

Ogbu, J. (1989). The individual in collective adaptation: A framework for focusing on academic under performance and dropping out among involuntary minorities, found in *Dropouts from school: Issues, dilemmas and solutions.* Albany, NY: SUNY Press.

Ogbu, J. (1987). Variability in minority school performance: A problem in search of an explanation, *Anthropology and Education Quarterly, 18*, 312-334.

Peña, R. (1995). Teacher expectations and students with challenging behaviors, *Current Issues in Middle Level Education, 4*, 67-78.

Peña, R. (1994). *Organizational uncertainty: Public schools and students with challenging behaviors.* Unpublished doctoral dissertation, University of Wisconsin-Madison, WI.

Raudenbush, S. & Bryk, A.S. (1986). A hierarchical model for studying school effects, *Sociology of Education, 59*, 1: 1- 17.

Raudenbush, S. & Bryk, A.S. (1989). The effects of high school organization on dropping out: An exploratory investigation, *American Educational Research Journal, 26*, 3: 353-383.

Rosenholtz, S.J. (1989). Workplace conditions that affect teacher quality and commitment: Implications for teacher induction programs, *Elementary School Journal, 89*, 4: 421-439.

Scheurich, J.J. & Imber, M. (1991). Educational reforms can reproduce societal inequities: A case study, *Educational Administration Quarterly, 27*, 297-320.

Scott, J.W. (1988). Deconstructing equality versus difference: Or, the uses of poststructuralist theory for feminism. *Theory, Curriculum and Society, 10*, 4: 127-142.

Spring, J. (1976). *The sorting machine: National educational policy since 1945.* NY: David McKay Company Inc.

Strauss, A. (1987). *Qualitative analysis for social scientists.* New York: Cambridge University Press.

Taylor, S.J. & Bogdan, R. (1984). *Introduction to qualitative research methods: The search for meanings.* 2nd ed., New York: John Wiley and Sons.

Turnbull, A.P. & Turnbull, H.R. (1990). *Families, professionals, and exceptionality: A Special partnership.* 2nd ed., NY: Macmillan Publishing Co.

Weber, M. (1947). *The theory of social and economic organization.* NY: The Free Press.

Wehlage, G.G., Rutter, R.A., Smith, G.A., Lesko, N., & Fernandez, R.R. (1989). *Reducing the risk: Schools as communities of support.* Philadelphia, PA: Falmer Press.

Wehlage, G.G. & Rutter, R.A. (1986). Dropping out: How do schools contribute to the problem? *Teachers College Record, 87*, 374-392.

Willis, P. (1977). *Learning to labour: How working class kids get working class jobs*. London: Saxon House.

About the Author

Robert A. Peña

Robert Peña is an Assistant Professor in the Division of Educational Leadership & Policy Studies at Arizona State University. He was graduated in 1993 from the University of Wisconsin at Madison with is Ph.D. in Educational Administration. Dr. Peña was awarded his M.A. in Secondary English Education in 1988 from SUNY Buffalo and his B.S. in English in 1984 from Buffalo State College. He has served as a public school principal, assistant principal, program coordinator and teacher.

Dr. Peña's research interests include poverty, interagency collaboration and their implications for value centered leadership. He also writes on the relationship between organizational theory and behavior and disadvantaged students and families.

Email: rpena@imap2.asu.edu

Copyright 1997 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://olam.ed.asu.edu/epaa

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey@olam.ed.asu.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

John Covaleskie
Northern Michigan University

Andrew Coulson
a_coulson@msn.com

Alan Davis
University of Colorado, Denver

Sherman Dorn
University of South Florida

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
hmwkhelp@scott.net

Thomas F. Green
Syracuse University

Alison I. Griffith
York University

Arlen Gullickson
Western Michigan University

Ernest R. House
University of Colorado

Aimee Howley
Marshall University

Craig B. Howley
Appalachia Educational Laboratory

William Hunter
University of Calgary

Richard M. Jaeger
University of North Carolina--Greensboro

Daniel Kallós
Umeå University

Benjamin Levin
University of Manitoba

Thomas Mauhs-Pugh
Rocky Mountain College

Dewayne Matthews
Western Interstate Commission for Higher Education

William McInerney
Purdue University

Mary P. McKeown
Arizona Board of Regents

Les McLean
University of Toronto

Susan Bobbitt Nolen
University of Washington

Anne L. Pemberton
apembert@pen.k12.va.us

Hugh G. Petrie
SUNY Buffalo

Richard C. Richardson
Arizona State University

Anthony G. Rud Jr.
Purdue University

Dennis Sayers
University of California at Davis

Jay D. Scribner
University of Texas at Austin

Michael Scriven
scriven@aol.com

Robert E. Stake
University of Illinois--UC

Robert Stonehill
U.S. Department of Education

Robert T. Stout
Arizona State University

---

archives | abstracts | editors | board | submit | comment | subscribe | search

EPAA • enter the archives • browse the abstracts • the editors • the edit board
• submit article • submit commentary • search • subscribe
volume: • 1 • 2 • 3 • 4 • 5 • 6 • 7

This article has been retrieved 3343 times since April 16, 1997.

Education Policy Analysis Archives

Volume 5 Number 11 April 16, 1997 ISSN 1068-2341

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass Glass@ASU.EDU.

College of Education

Arizona State University,Tempe AZ 85287-2411

Copyright 1997, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

Centralized Goal Formation and Systemic Reform: Reflections on Liberty, Localism and Pluralism

Kenneth A. Strike
Cornell University

Abstract

This paper asks whether there are reasonable concerns about liberty raised by standards driven systemic reform. **Part I** explores three kinds of concerns, students' interests in autonomy and authenticity, academic freedom, and pluralism. **Part II** explores two ways of conceptualizing the balance between liberty and various public interests, neo-classical economics and contemporary conservative thought. The paper draws two major conclusions about standards driven systemic reform: (1) This picture of reform raises serious questions about liberty. It may be inconsistent with some liberty interests of students. It is likely to pose serious questions about academic freedom and about pluralism. These concerns should make us cautious about systemic reform and should motivate us to a broader discussion of its assumptions and consequences. (2) The best defense of public sector reform efforts against their market oriented competition is one that emphasizes the importance of political goods such as citizenship. However, standards driven reform needs to avoid linkage with any nationalistic form of communitarianism. In order to do this it needs to seek ways to balance the demands for centralized goals and an educational system with an equal concern for local democracy, pluralism and community. A view of standards and accountability that is narrowly focused on clear public interests is crucial.

The paper concludes with an argument that we need to focus attention on the question of what makes for good educational communities, a discussion that is not abetted by debating issues of reform in a framework that poses choices between public sector and market approaches.

Standards driven, systemic reform, as envisioned in the national standards of Goals 2000 (1993) or various state reform efforts, involves centralized goal formation. Centralized goal formation requires developing detailed descriptions of curricula at the state or national level. Systemic reform aspires in addition to create an educational system in which various components are aligned so as to interact in a mutually reinforcing way.

Centralized goal formation may run afoul of certain liberties beginning with the liberty of students or their parents to be free from unreasonable educational coercion, but rising to problems of multiculturalism, parent's rights, freedom of religion, local control, federalism, and academic freedom. Linking centralized goal formation to the aspiration to develop an aligned educational system also raises the prospect of additional centralization of educational authority.

These tensions have not been lost on advocates. Two solutions have been proposed. One is "centralized localism" in which standards developed at the national or state level can be further specified in local jurisdictions. In their paper on systemic educational reform, O'Day and Smith (1993) suggest a view of "content-driven systemic reform" which "would marry the vision and guidance provided by coherent, integrated, centralized education policies common in many nations with the high degree of local responsibility and control demanded by U.S. tradition." (p. 128) (Also see Clune, 1993.)

A related strategy involves centralizing goal formation while decentralizing implementation. In the introduction to their recent collection on reform Finn and Walberg (1994) say "In this volume, Albert Shanker ... shows how outcomes-based education might best be employed within a national system of goals and a national assessment of progress." (pp. xxii) This passage is immediately followed by one commending "... the transfer of political power ... from producers to consumers (the choice movement) or from central controllers to decentralized decision makers..." (p. xii) The State of New York has also expressed a simultaneous enthusiasm for centralization and decentralization in its *A New Compact for Learning* (1994).

> ...this is the essence of the New Compact for Learning: that schools and school districts exercise initiative to make what changes may be needed to bring about the learning results we all desire. In a new relationship between the State and localities, the State defines more precisely what is to be learned, and local teachers, administrators, and boards of education have more freedom to decide how such learning is to occur. (p. 10)

This mix of ideas hang together as a view *if* we see standards driven, systemic reform as the intersection of three concerns. The first is the desire for high standards. The second is the desire for a coherent or "aligned" educational system. The third is the desire for a less bureaucratized system (Fuhrman, 1993). The first two concerns account for the emphasis on centralized goal formation. A national or a statewide curriculum is seen as both a means to excellence and a precondition of a system. Permitting local elaboration of state or national goals and localizing responsibility for implementation serves the third concern.

These are plausible aspirations. Yet it is not self evident that we can satisfy all three simultaneously. The idea that centralized goal formation can be reconciled with local autonomy by permitting local specification of centrally achieved standards is not easily reconciled with the idea that centralized goal formation will be the centerpiece in creating an aligned system. Arguably, the more local variation we have, the less of a system we will have. O'Day and Smith (1993, p. 297) suggest that ecology might be taught differently to students depending on their locale. Students in fishing communities might study coastal ecologies while students in Arizona might study deserts. However, the more we imagine this, the less we can imagine an aligned system. Is preservice teacher education going to emphasize the ecology of deserts or of coasts or just principles of ecology? Accountability is particularly a problem. It is easy to write standards capable of local variations in implementation so long as they are not too detailed. But tests are the operational definitions and enforcers of a curriculum. Tests which are affordable, provide for comparability, and respect local variation are at best difficult. (See Smith, Scoll, and Link, 1995 where the problem is recognized.) Is New York State going to have different science exams for students on Long Island and in the Adirondacks? Or suppose that some schools wish to emphasize a curriculum in which biological theory is central, others a curriculum in which ecology is emphasized, and a third group wish to focus on a tech prep agriculture curriculum? These are all responsible emphases in the biological sciences. All might be done with excellence. Each might be responsive to a local need. To the degree that we permit them, we will find it hard to align a system. To the degree we forbid them we will not have local autonomy. Thus a central question in thinking through systemic reform is to decide how much the desire for a coherent system will need to be accommodated to desirable forms of localism or pluralism.

Nor is it clear that the "the state determines the what and the locale determines the how" story can be consistently maintained. If systemic reform is to mean anything, it must mean that the state will be involved in such activities as helping or coercing failing schools to change (perhaps even "reconstituting" some), in exporting best practice, or enforcing opportunity to learn standards. While such activities may be commendable, it is not at all clear that they are consistent with the rhetoric of the decentralization of responsibility.

The issues run deeper. While the motivation for centralized goal formation, an aligned educational system, and debureaucratization reflect a concern for efficient educational organizations, the forms of decentralization that characterize American polity are often motivated by a concern for liberty. This is one point of federalism. The division of powers and the numerous jurisdictions that characterize American politics were created, in part, to prevent one faction or one idea from becoming the ruling faction or idea. Similarly, the autonomy of intellectual professions is intended to protect intellectual freedom. There are numerous examples where liberty has been asserted as an interest inconsistent with centralized authority over education. In *Pierce v Society of Sisters* the Supreme Court said that

> The fundamental theory of liberty upon which all governments in this Union repose excludes any general power of the State to standardize its children by forcing them to accept instruction from public teachers only. The child is not the mere creature of the State; those who nurture him and direct his destiny have the right ... to recognize and prepare him for additional obligations." (p. 12)

Pierce echoes a famous statement by J. S. Mill (1956) in *On Liberty*.

> All that has been said of the importance of individuality of character, and diversity in opinions and modes of conduct, involves ... diversity of education. A general state education is a mere contrivance for molding people to be exactly like one another; and as the mold in which it casts them is that

> which pleases the predominant power in government ... it establishes a despotism over the mind.... (p. 129)

In *Milliken v Bradley* and *San Antonio v Rodriguez*, the Supreme Court claimed that local control of education had Constitutional standing in that it served the liberty interest of permitting a higher level of control over the goals of education. New York's Court of Appeals asserted a similar view in *Levittown v Nyquist*. Law Professor Stephen Arons (1994) declared a national curriculum to be "broadly inconsistent with the principles of constitutional democracy." (p. 57) Others have seen standards as inconsistent with the freedom of parents or children. Professor Nel Noddings (1992) associates national goals with an ideology of control and as inconsistent with an ethic of caring.

Note the diversity and complexity of these contentions. Much of the objection to Goals 2000 has been in the name of federalism or local control. And the supposition that federalism and local control provide protection for certain forms of liberty is plausible. However, as we should have learned from the history of desegregation, sometimes local control permits local majorities to oppress local minorities. Local control is a significant factor in fiscal inequality. Moreover, the assumption that links liberty to local control is that the important forms of diversity are mapped onto geography so that providing for the autonomy of some political jurisdiction against the center promotes the liberty of those who reside in there. However, many of the forms of diversity that are important to Americans, ethnic or religious diversity for example, do not map onto local political jurisdictions in this way. Thus, even if it is true that centralization of authority over curricular content at the federal level is inconsistent with liberty or diversity, it does not follow that federalism and local control are always the solutions. Solutions depend on the kinds of liberty and pluralism that are desirable. While this paper will worry much about liberty and pluralism, it is not simply an argument for federalism and local control.

Liberty is not absolute. We are not free to run red lights, throw our trash into the streets, or install ourselves in public office by force. Even in the freest of societies individual conduct is appropriately compelled or restrained in order to protect the rights of others or to promote the public interest. Nor are all liberties and public interests of equal weight. A child who is not free to leave school because he or she is bored might be free to do so if he or she wishes to observe a religious holiday. We might approach the question of whether there are liberty interest that are at odds with centralized goal formation and systemic reform as follows: First, we need a sample of relevant liberties. Second, we need a view of the public interests that might be balanced against these liberties. Finally, we need some principled ways to weight these interests so as to strike a principled balance.

In **Part I**, I shall focus on developing a sample of liberty interests. I shall address three questions:

1. Are there any liberty interest of students that might be in conflict with systemic reform?

2. Are there any liberty interests of educational professionals or of the academic professions that are in conflict with systemic reform?
3. Is systemic reform inconsistent with diversity or pluralism?

In **Part II** I shall try to construct two different and more systematic way of understanding the nature of both liberty interests and public interests, and I will discuss some of the theoretical and political issues raised in attempting to balance them.

Part I

Students

One might argue that in a free society even modest interference with the educational preferences of students requires justification. Justification is easily provided. Education serves numerous public interests including human capital development and citizenship that are of sufficient importance to warrant constraint on the educational freedom of minors. The immaturity of students provides a further reason for adult direction of their education.

This picture might change if there are plausible views of education that serve important liberty interests. Are there such interests? Two candidates might be authenticity, the ability to choose and act in accordance with one's identity or nature, and autonomy, the capacity to freely and wisely choose among options unencumbered by prejudice or ignorance. Moreover, there are plausible educational programs associated with authenticity and autonomy.

Recall that Goals 2000 emphasizes the mastery of a common academic curriculum. Goal #3 of Goals 2000 says "By the year 2000, all students will leave grades 4, 8, and 12 having demonstrated competence over challenging subject matter including English, mathematics, science, foreign languages, civics and government, economics, arts, history, and geography..." Second, the reform movement has understood the point of such a curriculum as economic competitiveness and the development of human capital. (Hanushek, 1995; National Commission on Excellence in Education, 1983; Reich, 1990.)

No curriculum is likely to do much for students who ignore it. (See Steinberg, 1996.) Thus some school reformers have wanted to tell a story about how students are to be engaged in learning. One story goes like this: Students are only likely to be motivated to learn when something they want is contingent on their performance. Currently little that students value is contingent on their doing well in school. Indeed, as a society, we absurdly protect students from educational malfeasance by numerous second chance institutions. However, once we, as a society, have been able to agree on what students are to learn, we will be able to link school achievement to a range of incentives. Jobs and further education are favorite examples. (See Bishop, 1989 and Shanker, 1994.) The educational system is to speak softly and carry a big carrot.

Does such an approach interfere with any significant liberty? Return to an earlier Presidential report, *Youth: Transition to Adulthood* (Coleman, 1974). This work is noteworthy because, while it sought to change American education, it had a remarkably different picture of the problem to be solved. If the problem James Coleman (the committee chair) saw could be put into a phrase, it might be "youth alienation." Schools are age segregated. They abet the development of youth culture. Schools substitute formal for participatory learning and make the lives of adults and their communities invisible. Schools provide only one way to grow up.

The idea that schools provide too few ways to grow up might serve as a focus of discussion. Howard Gardner's (1982, 1983) work on multiple intelligences suggests one way of formulating the issue. Gardner claims that intelligence is not one thing but many. Schools focus on the development of a narrow range of these intelligences. Even if one disagrees with Gardner's formulation, it is apparent that schools make some capacities and aspirations central and others less so. In doing so they influence how students are valued by schools, whose capacities count for something and whose don't, and whose interests are worthy and whose aren't.

Schools cannot be about everything. Nor is it clear that what they are about should be fully determined by students' capacities or interests as much as, perhaps, social or economic needs or students' long term needs. Nevertheless, it may be harmful and alienating to students to subject them gratuitously to a curriculum that emphasizes capacities and aspirations that they do not possess in abundance. Moreover, often our sense of what we are good at or what we aspire to is part of our sense of who we are. When an institution does not value something central to our sense of self, we do not belong. We cannot generate a sense of ownership. We are alienated. We will resist or move through passively. If so, it is not a trivial thing if we develop a curriculum that gratuitously narrows the range of things that are valued in schools. Yet it seems that is a likely consequence of the strong focus of systemic reform on academic learning.

What does this have to do with the liberty interests of students? Perhaps, one kind of liberty is the freedom to be who we are, the freedom to act in accordance with our nature, identity, or character. Let's call this authenticity (Taylor, 1991). Honoring authenticity might include providing the freedom to pursue a kind of learning in tune with how we learn, what we are good at, how we understand our identity, or what we aspire to. One way to cash this out is to claim that schools ought to help students form an educational project that is both worthwhile and is their own. (See Noddings, 1992 for a useful example.) Such an educational program would not simply permit students to follow their interests. It would constrain or require what students pursue both for their own welfare and the public interest. And it would insist that students do something of worth. However, it would seek to honor their self-chosen educational projects within these limits.

Another perspective can be developed by considering the anomalous treatment of a liberal arts curriculum. The current emphasis on academic learning is linked to the concern for economic competitiveness and human capital formation. Yet this instrumental justification is quite different from arguments historically made for the liberal arts which have emphasized their intrinsic worth and have seen their chief values as including their capacity to transform the individual and to liberate the intellect (Adler, 1982). This liberation might be thought of as an interest in autonomy.

Such an education aspires to transform what people want. People are to come to want what is genuinely good. Their wants are to be consistent with virtue and self rule. Since this kind of education aspires to transform wants, that students who are initially exposed to a liberal education do not immediately aspire to its ends is not decisive. However, freedom in their education may be respected in that a liberal education should be such as to produce people who are virtuous, of good taste, and liberated intellect, and such people should come to view the education they have received as the kind of education they would have chosen for themselves had they been the kinds of people they now are.

Authenticity and autonomy are examples of liberty interests of students that are associated with plausible educational programs. How are such aspirations served by the curriculum envisioned by current reform efforts? Authenticity fairs poorly. Because systemic reform promotes a common academic curriculum, it is likely to make the "only one way to grow up" problem worse. It is unlikely to invest much effort in helping students to develop and pursue their own educational projects. If Coleman's arguments have merit, we might expect systemic reform to increase the alienation and disengagement of some students.

It is less clear that the liberal arts picture of the liberty interests of students is frustrated. Often those who accept the ideals of liberal education have thought that they are served by a common academic curriculum. Even so, I suspect these aspirations are not likely to be well served by systemic reform. Standards are to be a means of attaching school achievement to incentives such as further education and jobs. Students are thus pictured as people who value income and increased economic opportunity rather than as people capable of an intrinsic interest in history or science. National goals are rarely described as ways to alter students' educational preferences in order to help them love learning. There is no Dead Poets Society here, no sense that education can or should transform people. Anomolously, given the traditional values of the liberal arts, the motivation to learn is instrumentally attached to untransformed wants.

We need to be careful about such pictures. They may be more than empirical observations. They may presuppose a picture of human life that has been characterized as possessive individualism (Young, 1990). When schools make such assumptions about what students want they may abet a process of producing people with these wants. At the very least they may uncritically abandon the project of helping students acquire an enlightened set of tastes. They may privilege one conception of a good life and one picture of the point of education over another. It is a serious matter if we are implicitly defining the character of a good life as money, jobs, and things and transmitting such a picture to students.

Thus centralized goal formation and systemic reform do not comfortably fit with either authenticity or the liberal education of students.

Academic freedom

At first glance centralized goal formation with local responsibility for delivery seems consonant with the aspiration of teachers to be professionals. Teachers might expect more flexibility and autonomy to result from locating the responsibility for implementation at the local level. Nevertheless, teachers may find that the essential message of this formula is captured by the expression "You are free to do what we say in any way you choose." This falls somewhat short of a motto for empowerment. Teachers might want to claim that part of their professional expertise is knowledge of what constitutes excellence in their fields and that a process that defines goals politically at the state or national level, even if teachers participate significantly in the process, is inconsistent with their professionalism. (See Wise, 1990.)

One point of academic freedom is to protect the autonomy of the intellectual professions. The autonomy of intellectual professions serves an important public interest. A society cannot have a free intellectual life in which important matters are dealt with on the basis of reason and evidence if the intellectual professions are subject to unreasonable political control. However, the autonomy of the intellectual professions needs to be balanced against the democratic control of education. The character of this balance is difficult to describe. (See Strike, 1990.) Here I will only suggest that a line that should not be crossed is crossed when political processes come to determine what is regarded as true within disciplines or when political means are employed to make decisions that require significant expertise. However, when the question is primarily one of balancing conflicting interests or competing values, then political processes are more appropriate. Thus the decision as to whether to teach biology or auto mechanics calls for a political judgment. However, the question of whether the theory of evolution is true is a question for biologists. Balancing interests is a political matter. The truth value of biological theories is not. Roughly this suggests a division of labor in which the broad outline of the curriculum to be taught in public schools is a matter for political determination, but where the content to be taught within these broad outlines is appropriately left to experts and to processes of academic argumentation. It is thus worth asking if centralized goal formation and systemic reform might tend to politicize matters appropriately left to the disciplines or in some way degrade the working of the marketplace of ideas.

Centralized goal formation will not result in the content of the curriculum of public education being developed by politicians, lobbyists, and bureaucrats drafting standards in smoke filled rooms and taking votes on the laws of physics. The production of goals and standards is usually farmed out to teachers and academics. This process of goal formation may have more intellectual integrity than one in which the commercial interests of textbook companies are central. Nevertheless, national or state wide curriculum standards may tend to make curricular content highly visible and accessible to political influence. When standards are also controversial, the result is that they can easily move into the political arena. Recent debates about the American history standards should serve to illustrate. (See Manzo, 1997.)

History tells a story. It should have a narrative structure, a story line that brings its elements into an integrated whole. Two common narrative structures for American history are these: One is a narrative that emphasizes the spread of Western civilization. This is picture of the expansion of civilization, the triumph of Enlightenment culture, and of carving a nation out of the wilderness. An alternative is to organize history around an oppression narrative. Such a narrative might describe European emigration as an invasion. Slavery will not be seen as a deplorable, but corrected, aberration, but as an expression of the racist character of America. Our history is not a matter of working out the kinks in a more or less sound design, but is an ongoing liberation struggles against an oppressive order. One of the indictments of the Goal 2000 history standards saw them as dominated by an oppression narrative. (Fonte, 1995.)

A history curriculum might weave together various narratives or even note the competition between them. There is no obligation to have only one. But having no narrative is not an option if we wish history to make sense. Apart from some narrative structure history becomes "one damn thing after another." It becomes impossible to explain why some facts are included and others not and impossible to interpret what is included. The standards for world history have been widely criticized for just such defects. They are very long, over 300 pages with over 526 standards, yet there seems no explanation of why the Scythians, the Xiognu, and the Olmecs receive more attention than the ideas of Islam or Protestantism (Gagnon, 1995). Thus a dilemma. History is not just a series of facts. It requires a narrative. But any narrative will be controversial. How, then, are we to have standards that are not vacuous or platitudinous without giving some controversial conception of the American story the force of law?

This point about history can be extended to other subjects. Academic disciplines are frequently characterized by competing pictures of the nature of the field. People will and should structure curricula so as to reflect their understanding of the nature of their field. Biology has traditionally been structured around taxonomy. Natural selection and biochemistry are now more central. The post Sputnik structure of knowledge movement made set theory central in mathematics instruction. That emphasis has largely disappeared in the NCTM standards and is replaced by something called constructivism (Mathematics, 1989). Thus even in mathematics which, at the K through 12 level, is not characterized by disputes about what is mathematically true, there will be debate about matters important to teaching. Whole language v phonics is a similarly contentious matter.

Goals 2000, systemic reform, and the development of standards are not responsible for controversies about narrative structures, pedagogical approaches, or the organizing characteristics of disciplines. However, they contribute to the fact that these debates have begun to take place on the floor of Congress and in state legislatures. The criticism of state and federal history standards is largely owing to the fact that the standards are seen as rooted in the oppression and struggle narrative. Those who have attacked them prefer the spread of Western civilization narrative. Former under secretary of education Chester Finn (1996), has suggested that bus drivers, policemen, shopkeepers, engineers, preachers, and orthodontists should have a larger role is developing standards than historians. (This "democratization" of standards makes sense only if one assumes that the truth of historical claims is subservient to the political or socialization role of history so that, perhaps, Americans are entitled to "feel good" history regardless of its truth.) Similarly, the development of standards is not responsible for the emergence of constructivism or for the existence of state textbook selection committees, but it is responsible for the fact that state textbook committees have become an enforcement mechanism for this doctrine (Saxon, 1995, 1996).

We are now in a position to consider four possibly undesirable consequences of goal centralization of the sort required by systemic reform.

First, developing centralized standards places controversial academic issues on the public agenda in ways that abet their politicization and transforms them from a search for truth into a search for the politically acceptable. The history standards are exhibit A. Language standards may pose similar questions given a need to resolve issues about the status of phonics or standard English (Gagnon, 1995) or likely debates about multiculturalism. Math may provoke a struggle between those who emphasize practice and the currently popular constructivism (Saxon, 1995). Biology standards may put evolution on political center stage. Systemic reform is not responsible for the fact that these matters are controversial, but it abets a process which makes them visible, takes them out of the hands of teachers and members of the academic professions, and puts them into a political arena. At worst, it may result in disputed doctrine having the force of law.

Second, systemic reform provides paths for interference in the internal workings of universities. Part of creating an articulated system has to do with teacher education. O'Day and Smith (1993) suggest that states will need to work with universities to ensure an alignment between the training of teachers and new curricular frameworks. We should wonder about the form this working with will take and take careful note that systemic reform seems to include peoples' professional opinions as among the things that need to be aligned. Will history professors be expected to take the view of history expressed by the history standards when they teach future history standards? Will math professors be expected to be constructivists? Probably not. Will people who train teachers? This seems more likely. Can we assign any meaning to alignment that does not involve someone external to the university generating expectations that define for scholars (and pursuantly teachers) how they are to understand the nature of their subjects or who will be hired or whose work will be funded? I have difficulty in constructing a picture of this alignment that does not involve some politically achieved picture of a subject matter being made normative for a portion of the professorate. If something like this does not happen, there is no systemic reform of teacher education.

Third, centralized goal formation may generate pressures tending to prematurely decide open questions. Coherent standards and curricular frameworks require some view of the organizing principles of a field. These are ordinarily controversial and fleeting. The post WWII period has seen a number of views concerning how to structure curricula. We have swung from a focus on behavioral objectives and the Skinnerian conviction that complex skills could be constructed from their parts, to the top down view of the structure of knowledge movement, to the current emphasis on constructivist learning. Views about the basic nature of disciplines or their pedagogical organization have always been controversial, yet believers have often been confident enough to be willing to use the power of government to impose their vision before they have won the academic argument. Standards driven systemic reform is an excellent mechanism for getting the profession out ahead of the evidence and for quieting contrary voices.

One thing that protects the marketplace of ideas is the systemic incoherence of our educational system. Systemic reform may provide a tool for eroding this protection, for enhancing the systemic implementation of the currently dominate fad, and for transferring power over teacher education from the professorate to state education personnel and their academic allies. Once again we are asked to place our faith in those intellectual elites who have promised progress in schooling through scientific research and management. (See Tyack and Cuban, 1995.) Yet no one with a sense of the recent history of educational fads and reforms can possibly doubt that in another decade current views will be represented as the problem, not the solution. Other visions of reform will be ascendent. To base current policy on the conviction that educational research has given us the assurance that we can now make policy systemically without the fear that we will write our errors large is naive. It is part of the human predicament that we must often act in the face of uncertainty. We cannot stop educating pending definitive research. But we do not need to act in concert in the face of uncertainty. Uncertainty is a reason for decentralization, for letting a thousand flowers bloom. J.S. Mill (1956) reminded us in *On Liberty* that when we enforce our opinions using the power of the state, we assume our infallibility. Let us not.

The fourth difficulty is ossification. The previous objection would loose some of its force if we could believe that the development of standards would not curtail the debate about what standards should be and that standards would be kept up to date as opinion changed or knowledge grew. Such a view defies experience. Once in place current standards will be difficult to root out. They will have the force of institutional inertia behind them. Texts will have been written, tests made, lessons planned, money spent, appointments made. The current vision will gain a constituency. Meanwhile, governments will change, new concerns will emerge, the attention of reformers will move elsewhere. National goals and state curriculum standards will have the same degree of flexibility as, say, New York State's Regents exams, which is to say, some, but not much. What we manage to put into place may live to be dated and will be seen by a later generation as something that ties teachers to an outdated view of their subjects.

These four dangers are dangers of trying to write good and informative standards. There is another option. We may prefer consensus to coherence thereby running the danger of standards that are vacuous and platitudinous. Such outcomes are likely in controversial areas. They are even more likely when the groups that produce the standards are broadly representative as they should be. Such groups are not suitable forums for resolving complex intellectual disputes. Members of such groups are likely to approach their tasks by seeking accommodation, not by pursuing the argument until someone wins the debate.

There are many questions of public policy that are usefully approached by seeking accommodation and consensus. Questions about where to locate highways or how to fund public works are examples. Moreover, since highways must go somewhere seeking consensus is unlikely to produce vacuous results. Vacuity is more likely when the task given to a diverse group is to produce a piece of paper. New York's Preliminary Draft of a Framework for Social Studies (1995) often seems to succumb to the vagaries of consensus. Are we to emphasize depth or breadth in our coverage? Shall we emphasize unity or diversity in our understanding of whether Americans are a people? Predictably, in each case, the answer is both. We must seek an appropriate balance. The guidance as to how we are to strike such a balance is meager and unlikely to resolve many disagreements. Or consider Standard 6.

> Students will use a variety of intellectual skills to demonstrate their understanding of the basic civic values of American constitutional democracy; the roles, rights, and responsibilities of citizenship; and the avenues of participation in American life. (p.13)

The performance indicators that follow provide such guidance as that students should know something about the Constitution and about how citizens can influence policy. Who will disagree? Will people who have very different ideas about the character of the U.S. Constitution or citizen participation, or who have conflicting views about how these topics should be taught find much to quarrel with here? If they do not, then what guidance is provided? Will there be many social studies teachers who do not currently believe themselves to be doing these things? What current educational practices are inconsistent with this?

When standards are vacuous academic freedom is still an issue because the development of standards and frameworks can effect a change of venue with respect to where and how educational questions are discussed. Issues are moved from academic forums where the truth of various ideas is argued about to political forums where consensus is sought. This change of venue is not necessarily a good thing. It will not do the intellectual quality of educational debate much good to pose difficult and contentious issues in consensual forums. Moreover, it invites a process where the real standards are developed by test makers working out of the public eye.

The centralization of goal formation thus creates a dilemma. It creates political forums which need to decide contentious and open issues. It may produce results in which currently popular doctrines becomes ruling doctrines and minority views are subordinated. Or it may generate consensual processes in which disagreement is buried beneath platitudes. In both cases the free and open exchange of ideas and the process of criticism and debate is truncated.

Pluralism

My final concern is with pluralism. Pluralism might be at odds with centralized goal formation and systemic reform in three ways. First, there might be conflicts over content with claims that national or state wide frameworks promote cultural imperialism. Second, there might be issues of opportunity with claims that the mandated curriculum makes demands of time and resources of sufficient scope so as to preclude the opportunity to pursue studies of particular interest to some group or locale. This might be particularly burdensome to minorities to the extent that they bear higher "opportunity costs" for cultural reproduction (Kymlicka, 1989). Third is the issue of process, where it might be claimed that the centralization of goal formation takes certain discussions out of local communities and locates them at the state and national level. This diminishment of opportunity for participation might be viewed as inconsistent with the community constituting function of local deliberations.

To pursue these concerns I want to sketch the position of Harvard philosopher John Rawls (Rawls, 1971, 1993). For Rawls a central question of political philosophy is how we can have a shared view of justice, one consistent with the view that we are all free and equal, but which does not unreasonably constrain our ability to have and pursue our own vision of a good life. Part of Rawls's answer is that our shared view of justice should be philosophically shallow in that it does not presuppose religious or philosophical conceptions not widely shared. Rawls also emphasizes such liberties as freedom of conscience and freedom of association. People are entitled to form their own conception of their own good and to share and pursue it with like minded others. The state is expected to show neutrality between these competing visions of a good life. (For discussion see McCarthy, 1994; Okin, 1994; Scheffler, 1994.)

How might these ideas be applied to schooling? What may follow is that public schools may rightly seek to transmit to students a shared political culture rooted in a shared view of justice. They may thus promote citizenship appropriately conceived. They might also promote educational goals that are likely to be widely valued by people regardless of the vision of a good life they have. (See Strike, 1984.) However, they must seek to be neutral between different visions of the nature of good lives. Moreover, they may not understand the notion of justice or the character of civic education in such a way as to make second class citizens out of some because the vision of a good life of others is preferred.

Rawls's views place obstacles in the path of an extensive, common and legally mandated curriculum. It suggests that an obligatory common curriculum is consistent with a just and pluralistic society under one of three conditions. First schools might properly promote a common curriculum if it is necessary to for a just political culture. Second schools might promote a common curriculum that emphasizes "universal instrumentalities." Perhaps there are some knowledge and skills that are essential for the pursuit of one's vision of a good life whatever it is. Lliteracy is an example. Third, public schools might teach subject matter that is broadly uncontroversial given the range of differing conceptions of the good extent in society. Math is an example.

While the details of what is permissible given such criteria would be subject to much debate, it is certainly possible for reasonable people to doubt that the common academic curriculum envisioned by systemic reform can meet these criteria. They will be especially problematic in areas of controversy or areas replete with cultural diversity such as history, art or literature. Consider that while reformers have generally emphasized human capital formation as the purpose of reform, there has been a secondary claim that creating a shared curriculum is important for social stability, a democratic society, or equal opportunity (Hirsch, 1987). *A Nation at Risk* (National Commission on Excellence in Education, 1983), for example, claims that "A high level of shared education is essential to a free democratic society *and to the fostering of a common culture* that prides itself on pluralism and individual freedom." (p. 7) Or consider a passage from former Secretary of Education, William Bennett (1984).

> We are a product of Western civilization. That our society was founded upon such principles as justice, liberty, government with the consent of the governed, and equality under the law is the result of ideas descended directly from great epochs of Western civilization--Enlightenment England and France, Renaissance Florence, and Periclean Athens. These ideas, so revolutionary in their times yet so taken for granted now, are the glue that binds together our pluralistic nation....(p. 30)

These passages are ambiguous with respect to the concerns I have raised. They do salute some form of pluralism. Yet they also seem to suggest a political role for a broadly shared culture that is difficult to reconcile with the kind of pluralism Rawls emphasizes. This view looks to schools to secure political stability by providing the kind of shared learning experience necessary for Americans to be a common people. (See Schlesinger, 1992.) However, the areas of commonality required to produce a common people are likely to be those least likely to meet my Rawlsian criteria, history, art or literature. Rawls's view of justice suggests that these aspirations go beyond the reach of a liberal polity rooted in the search for reasonable pluralism. (For discussion see Kymlicka, 1995 and Strike, 1991c.)

It is unlikely that we can justify an obligatory common curriculum of the sort required by standards driven systemic reform on the grounds that it provides knowledge and skills that are either broadly accepted or are universal instrumentalities. Different life plans are likely to require different kinds of learning. Different cultural or religious groupings of citizens are likely to have different pictures of the kind of education suitable for the lives their members seek to live. African American parents may want their art, literature or history emphasized in the instruction of their children. Religious parents may be frustrated by an antiseptic history of America that seems embarrassed to note its religious roots. (See Vitz, 1986.) And farming communities, Adirondack residents, and coastal dwellers may want a different biology curriculum for reasons having nothing to do with what is true in biology. While it is difficult to see how any school system could accommodate every demand for diversity, it is also hard to understand why we should resist a priori the idea that the educational programs available to Americans should not differ in ways so as to reflect their diverse aspirations, histories, locales, cultures, religions and visions of good lives. We are only likely to reject such diversity out of hand if we tell ourselves that Americans need more than a shared political culture and that they should be a common people.

Let me once more return to some comments by O'Day and Smith (1993). In the concluding section of their article on systemic reform and equal opportunity they recognize that liberalism poses issues for systemic reform. They proffer a solution they call dialogical pluralism which they attribute to me. They see the discussion required to legitimate curricular frameworks as similar to the dialogue I hold is educationally important. They write

> An intriguing aspect of Strike's model of schooling is that it views schools and their students as undergoing the same learning process that we hope society will undergo as it works toward the goal of legitimate curriculum frameworks. The way that decisions are made about the content and pedagogical strategies of the frameworks is crucial to their legitimacy. There must be strong and continuing input from all the various stakeholders to ensure both the legitimacy of the content and the political buy-in of the stakeholders. (p. 297)

In the article to which they refer, I proposed a view in which both within group and between group dialogue is educationally important (Strike, 1991b). In the first case the point of dialogue is to deepen the understanding of one's own religion, philosophy or culture. In Rawls's language it is to develop one's capacity to have a conception of one's good. The emphasis on between group dialogue serves several purposes. One is the promotion of tolerance. The second is to subject one's own view to criticism, to learn from the other.

O'Day and Smith's views differ from mine in two respects. First, the forms of dialogue I emphasize are not aimed at consensus. Second, if they are to serve their purposes, these forms of dialogue need to occur locally. They must involve conversations between students, teachers, and parents (although employing resources from the broader community). They are aimed at learning and character formation, not decision making. They are not reasonably viewed as conversations between (the representatives of) stakeholders.

O'Day and Smith, however, view dialogue as a means to decision making at the state or national level. When decisions need to be made, dialogue of the sort they recommend is a commendable way to make them. But the prior issue is whether we ought to have a shared curriculum of the kind envisioned in Goals 2000. Consider an analogy. We might view the process of legitimization that O'Day and Smith advance as akin to an attempt to achieve a democratic consensus about the proper American religion. I am certainly an advocate of religious dialogue. I do not view it as a means towards religious consensus. Curricular consensus is appropriate when it is necessary to produce a shared and just political culture or when it promotes universal instrumentalities, but in a society committed to real diversity (on my Rawlsian view) democratic dialogue does not warrant the imposition of a common curriculum that exceeds the bounds of reasonable pluralism.

Systemic reform is predicated on the view that the efficiency of any organization depends on its ability to specify its goals clearly and to order its resources to seek its goals. This is a perfectly coherent view of how to run a business. It is less clear that it is a viable view of education. Some authors (Wise, 1990) have claimed that teaching is the kind of activity that is inherently difficult to regulate because it requires considerable judgment and flexibility. Such claims point towards teacher professionalization as the key to school reform. The argument I have made above is analogous to this view, but it takes a somewhat different path. It claims that good education is difficult to regulate in the way in which systemic reform seeks to do because of the way in which the idea of a good education is entwined with notions of human freedom.

In schools, people and ideas, not objects, are the "output." In a free society people may wish a say in their own education, and they may have rights not overcome by collective decision making. Their education may have to accommodate the public interest, but peoples' voices and their rights cannot be brushed aside to accommodate the requirements of an efficient system. Moreover, a good education is often more about the argument, the evidence, the debate than it is about a prespecified outcome. Participating in the argument is educative. The need for participation and to follow the inquiry where it leads rubs hard against the idea of an efficient system. Finally, we live in a society in which people have diverse religions, cultures, and visions of a good life. People may reasonably want schools to educate their children in ways sensitive to their differences. If so, we need to be careful when we seek to structure an educational system by beginning with politically defined outcomes of sweeping scope. Consensus building and democratic legitimization are not always adequate to warrant intrusions into these areas of difference. In a free society, sometimes liberty trumps both democracy and efficiency.

Part II

In the previous section I described several reasons to suppose that standards driven, systemic reform, may erode important liberties. Nevertheless, in a free and democratic society, liberty is not absolute. It must be balanced against public interests such as human capital formation, justice, political stability, and citizenship. In this section, I want to explore views about how liberty interests and their counterbalancing public interests might be conceived.

A "neo-classical" picture

Contemporary school reform is substantially motivated by the concern for productivity, international competitiveness and a changing job market. It thus seems reasonable to consider how liberty interests and public interest might be conceptualized within the framework of an economic view. Below I characterize a normative interpretation of a view that emphasizes the role of free markets. While I think the view will be recognizable as a version of neo-classical economics, its purpose is to conduct a thought experiment, not to rigorously describe a currently held position. Moreover, I am not so much concerned with whether this view is true of false as I am with what it makes possible to say or to argue. All theories provide a vocabulary and a conceptual framework for describing some range of human activities. Their conceptual resources may be more or less adequate for articulating diverse phenomena, and they can distort issues when concepts are generalized beyond their proper domains. What I suggest below is that the view I characterize suggests some things worth noting about the character of liberty interests and public interests, but that the framework is insufficient to represent all that is at stake.

A normatively constructed neo-classical account might claim the following:

1. People can be productively viewed as bundles of preferences.
2. Rationality can be understood as efficiency in maximizing the satisfaction of this bundle of preferences.
3. Other things being equal, markets are the most efficient way of maximizing aggregated preferences because they provide incentives for producers to provide what people want at the lowest cost.
4. The preferences that people have have normative force (Monk, 1990) in that, other things being equal, the mere fact that they are what is wanted is a reason why they, and not other preferences, ought to be satisfied. Markets are morally desirable as well as

efficient (Friedman, 1962) because they are responsive to peoples' actual preferences expressed through their choices.

5. Government interference with markets and consumer choice can be justified when markets fail to be efficient or when certain important values are not realized. Among the conditions that may warrant governmental interference are:
    a. When markets result in under investment in public or collective goods (such as human capital) because the benefits of such investment tend to accrue to people other than the investor (Friedman, 1962).

    b. When the results of market distributions are inconsistent with important political goods such as equal opportunity, justice, or political stability.

Centralized goal formation and systemic reform seem prima facie inconsistent with this picture and to a degree that exceeds more decentralized forms of public education. Most importantly, the goals of education are produced by highly centralized public deliberations instead of through consumer choice. Also, accountability is to the center, to the state, not to the consumer. Consumer choice, responsiveness to consumer preferences, markets, and market incentives are thus not much in evidence.

However, these difficulties might be overcome either if there is some good such as human capital development or equal opportunity that meet the criteria for justifying exceptions to the desirability of markets. I would propose (a thesis of this magnitude cannot be argued here), however, that neither of these interests is likely to render systemic reform plausible from the perspective of neo-classical economics.

A preliminary observation: the choice is between the view that education should be publicly financed but privately administered and the view that it should both be publicly financed and publicly administered. Neo-classical economists are likely to grant that public funding is required in order to rectify under investment in education and to promote equal opportunity. It is public administration that is suspect.

Evidence to substantiate the claim that systemic reform is more likely to promote human capital development than a market oriented approach is likely to be conjectural and controversial. (See Berliner, 1993 and Paris, 1995 for skeptical discussions of relevant empirical evidence.) But in policy when do we ever actually *know* something? We are almost always required to act on the basis of weak and conjectural evidence. This observation cuts two ways. On one hand, if the liberty interests at stake have only the force of "mere" preferences, decisive evidence may not be called for in order to over ride them. If we require the kind of evidence to overcome any preference of the moment that we should require to interfere with freedom of religion or some other liberty with constitutional standing, we will do nothing. On the other hand, a normal and reasonable response of academics to ambiguous evidence is to continue to affirm the basic assumptions of their "paradigm" or research program. (See Kuhn, 1970; Lakatos, 1970; and Strike, 1979.) They do this reasonably because a paradigm provides a fruitful picture of how the world works and because they believe that with additional research they can succeed in applying their view to new or recalcitrant cases. Since advocates of a neo- classical view are likely to be strongly committed to the view that markets in education (with public financing) ought to be more efficient than publicly administered schools, they are unlikely to be dissuaded by ambiguous contrary evidence. Thus, I doubt that those proponents of neo- classical economics who are otherwise predisposed to favor markets over the public sector will find that such evidence as exists for systemic reform provides sufficient reason to show that systemic reform is more likely to serve human capital formation than is a more market oriented system.

A second strategy to justify systemic reform from the perspective of neo-classical economics is to claim that public schooling serves desired political goods not served by markets. Among the political goals of schooling might be providing equal opportunity, creating the sort of shared culture required for a stable democratic society, or developing democratic character. (For discussion see Verstagen, 1991.) While no view of these political goods is strongly associated with neo-classical economics, its assumptions do constrain the view of justice that can be held.

There must be free labor markets. Thus, even where there is considerable leeway for redistribution of wealth, some income inequality resulting from differential possession of scarce talents or productive resources must be consistent with justice. Differences in acquired skills and ability to learn must be respected in hiring and in allocating educational resources.

Given this, neo-classical economics points to a view of justice in which equality of opportunity is a core conviction. Equal opportunity is consistent with a market economy and free labor markets and can be represented as a means to insure efficient use of human capacities by allocating them and the educational resources that develop them on criteria related to their efficient use. Equal opportunity will be viewed as fair competition (Strike, 1982).

This core belief in equal opportunity is easily made the basis of an argument for public funding of education. Apart from public funding the talents of those who cannot afford an education may go undeveloped and their competition with the more affluent will be unfair. However, here, as with human capital formation, it is not apparent that a theory of fair competition requires schools to be operated by the public sector. Such an view would need to show that public funding apart from public administration is inadequate to promote fair competition. While there are no doubt an abundance of arguments to examine here, what is important is that the conception of equality at issue is narrowly framed as fair competition in the labor market. This is the view of justice most likely to be served by the public financing private schooling. If the argument for markets is otherwise strong, ambiguous evidence about equal opportunity as fair labor market competition is unlikely to provide persuasive grounds for public sector operation of schools.

It might be proposed that decentralization of the delivery of educational services in the public sector simulates market delivery. This seems implausible. Market efficiencies depend on consumer choice and competition neither of which is approximated by public sector reforms such as site based management or parental participation in school governance. The plausibility of public sector decentralization as an educational reform depends on arguments that see bureaucratic control as a principal cause of educational failure or which seek reform through professionalism, parental involvement, and the democratization of schooling (Malen and Ogawa, 1990 and Strike, 1993). These arguments may be good arguments, but any association with market strategies is illusory. Systemic reform and an emphasis on market delivery of educational services represent competing views of how to secure effective education. Their common emphasis on decentralization should not be allowed to cloud this point.

Thus, if my conjectures are correct, neither the interest in human capital formation nor the interest in equal opportunity are likely to provide a very compelling justification for public sector operation of schools, let alone systemic reform, from the perspective of neo-classical economics. However, this argumentative weakness might be seen as largely a consequence of the fact that the public interests in education, human capital formation and citizenship, are narrowly framed so as to be those most likely to be served by publicly funding of private sector schooling. What follows is that the strongest justifications for public sector reform are likely to depend on or presuppose political goods such as tolerance or democratic citizenship that seem more likely to require common schools for their achievement. (For discussion of the link between citizenship and the public character of education see Callan, 1966; Gutmann, 1995; Macedo, 1995.) If such political goods were able to secure the need for public education, we might then claim that systemic reform was the most efficient and equitable public sector means to deliver educational services.

This discussion is suggestive of three conclusions:

1. Many of the concepts required for a full discussion of how to balance various liberty interests in education against various public interests are external to the framework of neo- classical economics. Within the confines of neo-classical economics we are likely to represent the interests to be balanced as the liberty interest in preferences on one hand versus human capital development and equality of opportunity on the other. These are relevant interests. But there are other liberty interests, freedom of conscience, intellectual freedom, or freedom of association, for example, that may be educationally significant, but which are not easily articulated via the concepts of neo-classical economics or in the language of markets and preferences. Similarly, there may be political goods such as democratic citizenship that need to be balanced against these liberties.

2. It may be that, as a view of efficiency in human capital development, systemic reform is justifiable only if we antecedently assume that education should be provided by the public sector. We are

only likely to make such an assumption if we appeal to political goods, such as democratic citizenship, broader than those easily characterized by the concepts of neo-classical economics. Here we should note an irony. Systemic reform emphasizes goals such as human capital formation, economic productivity, and international competitiveness. However, if I am right, its plausibility in its competition with market views of reform depends largely on its accepting broader goals such as democratic citizenship.

3. Once we have begun to wonder whether we need to consider a wider range of goals than human capital formation and fair competition in labor markets in our discussions of reform, we also need to be concerned with how we frame choices about reform. Much current debate about reform might be simplified as a response to the question "Which is the best approach to improve our competitiveness, the market or the state?" But if we need to consider a wider range of values in our deliberations, it may be that market versus state misrepresents our choices.

A useful illustration of this point is the debate about charter schools. The policy making community has tended to see the point of vouchers and charter schools as the creation of quasi- markets in education. However, the motivation of those who have sought to create charter schools and of the parents who have sent their children to them often seems more communitarian in character. That is, people start charter schools because they have a distinctive picture of the character of a good education that they are unable to pursue within the confines of most public schools. Thus they seek a less regulated environment and a community of the like minded. (See, for example, the description of the City on a Hill Charter School.) Here the tension isn't between market and state. It is between a view that wish to pursue its own goals and to emphasize freedom of association as the organizing principle of school communities versus those who believe that common schools organized geographically best serve democratic interests. These latter issues which I think are difficult to articulate in the language of neo-classical economics also seem to me to be the more important. And the question is "What is the nature of educative communities given the range of values that ought to be served?"

Finally, in what has proceeded I have taken it as obvious that human capital formation is a public good. However, this does not commit me to any view of that in which human capital ought to consist. Moreover, our conception of human capital also needs to be informed by our political goals. A liberal democratic society needs a view of human capital formation that encourages the virtues of liberal democratic citizens. If so, it will also need to reject the idea that the requirements of labor markets are the sole relevant consideration in forming a conception of human capital. Market versus state may not adequately frame the question of what should count as human capital.

In the next section of **Part II** I shall explore some ways to expand our conception of these more robust interests.

The conservative compromise

Modern conservatism might be interpreted as a merger between a rightist interpretation of neo-classical economics and cultural conservativism. The salient features of the first is an emphasis on consumer sovereignty and markets, a corresponding antipathy to the public sector, and a view of justice that is narrowly focused on equality of opportunity.

The basic idea of cultural conservatism is that the traditional values and practices of a society have prima facie normative force because they have met the test of time (and perhaps of reason or God) and because they are the social glue that forms social bonds and political community. Some of the commitments that might be held to flow from this are (1) The need to affirm the centrality of a culture whose roots are largely European as basic to national solidarity and democracy and a pursuant emphasis on a "Eurocentric" liberal arts curriculum; (2) a commitment to traditional virtues (honesty, moderation, courage, family values, etc.) as central to moral education; and (3) a commitment to the moral principles (if not the theology) of a broadly Judeo Christian morality.

The account I will construct of liberty interests depends on the insight that the privatization of schooling that many conservatives advocate potentially permits people to associate together for educational purposes in ways that reflect their religions, ethnicity or their particular and distinctive views of a good life. If so, then the liberties at stake might include freedom of religion, freedom of association, freedom of information, and the freedom to form and pursue one's own conception of one's own good. Several of these liberties have constitutional standing. Others are akin to them. Thus, the conservative's support of consumer choice provides support for a robust pluralism in education, a pluralism that is inconsistent with the agenda of cultural conservativism.

Recall that one source for the emphasis on a shared academic curriculum that has characterized the excellence movement is the cultural conservatism of such as William Bennett. The importance of centralized goal formation to the cultural conservative's agenda can be seen by considering the significant multicultural and, indeed, libertarian, potential that is otherwise inherent in the rhetoric of market decentralization when applied to schooling. If we are to take consumer sovereignty seriously, we are likely to view education as a process in which choice reigns restrained only by the rather modest notion of the public interest characterized above. Given this, it is hard to see why people ought not to be entitled to freely associate to pursue such conceptions of education as seem good to them. Why not religious schools, academies for black boys, gay schools, even coven schools, or, minimally, a robust elective system (charter schools, perhaps) in the public sector? Let a thousand flowers bloom. Is this not where consumer sovereignty leads?

But consumer sovereignty in education is constrained by the public interest in productivity and justice. Are we to believe that these public interests are sufficiently weighty to deny to people their right to an education of their choice even if the choices are some of the one's described above? There are certainly some interesting conversations that we might have about the coven school and, no doubt, about some of the others as well. However, the discussion of the neo-classical view suggests that it is implausible to claim that these forms of diversity are inconsistent with either the minimalist notion of justice that can be gotten from neo-classical economics or that the public interest in productivity precludes them.

Note the shift in the character of the liberty interests now involved. I have assumed that a due regard for consumer sovereignty would serve a larger range of liberty interests than simply allowing people freedom to pursue their own bundle of preferences. The examples noted above assume that given the freedom to do so people would form educational associations that reflect their common religions, ethnic backgrounds, or visions of the good. In doing so, I have begun to conceive the liberty interests involved in a way that makes them more important and thus less easily rebutted by claims about the need to develop human capital. The liberties involved now include freedom of association and freedom of religion. If such liberties are to be overcome by claims about the public interest, a more robust sense of what is involved in the public interest seems required.

Cultural conservatives have candidates for such interests. The following are possibilities. It might be argued, as Mr. Bennett does, that the political stability of American democracy requires us to affirm the centrality of European culture because it contains the roots of democracy. Political institutions are not easily maintained once divorced from their cultural roots. Moreover, any stable society requires significant elements of a common culture. People must understand one another to cooperate and to associate on amicable and equitable grounds (Hirsch, 1987). The degree of pluralism likely to result from the kind of application of consumer sovereignty envisioned above invites religious, ethnic, racial and cultural Balkanization.

Similarly, it might be argued that a commitment to traditional virtues (honesty, moderation, courage, family values, etc.) is central to moral education and to law, order and decency. Perhaps a shared Judeo Christian outlook is central to political stability, common understanding, and social order. (See Anthony Scalia's discussion of the importance of a civil religion in *Lee v Weismann*.)

What has changed? Essentially this: On one hand more robust claims for liberty in which the freedoms claimed are the freedom to associate and to pursue shared goals and group affiliations have been asserted. At stake now are our religions, our cultural identities, and our visions of a good life. The robust pluralism that potentially results has been met by a form of communitarianism that asserts the importance of the national community. This nationalist communitarianism has a more robust picture of the public interest involved in education. What is at stake is the stability and order of a democratic society. The public interest now links citizenship and virtue with a national culture. Conservatives have need of such public interests if they wish to constrain the plunge into diversity otherwise made available by their commitment to markets.

Both factions of the conservative compromise must give up something to participate in this vision. Market advocates must abandon the notion that preferences have moral force. In education, at least, they will need to see quasi-markets as means to more efficiently pursue agreed upon ends, not the means to pursue self-chosen ends. On the other hand, cultural conservatives need to believe that the cultural traditions they value can survive in a society otherwise dominated by the ethos of markets. (An implausible belief, I think. Cultural conservatives are not very astute as to who are their real enemies.) Both accommodations will be grudging and the resulting coalition unstable.

Against this background, centralized goal formation might be viewed as serving two purposes of cultural conservatives. First, it constrains the excessive curricular pluralism and Balkinization that is a possible consequence of a market emphasis in education. Decentralization will be a decentralization of means. We will view decentralization as debureaucratization or as the means of pedagogical experimentation. If there are to be vouchers or charter schools even these may be expected to serve national or statewide goals (Finn, 1996). We will see markets as a means to promote efficiency in education, but not as an encouragement to individualism or pluralism. Second, centralized goal formation sustains a shared culture. It seeks to ground schooling in common curriculum that reflects a Eurocentric, a North Atlantic, or even a distinctively American (Schlesinger, 1992) shared culture that can be the basis of a stable political culture, of common understandings, and of moral coherence.

Here cultural conservatism has triumphed over neo-classical economics in several ways. (1) Goal formation is left to the public sector. (2) Human beings are no longer looked at as bundles of preferences. They are now seen as culturally embedded - as formed and oriented to their lives by their religions, ethnicities, histories, and communities. However, the kind of communitarianism involved in cultural conservatism is nationalistic more than pluralistic.

In the American context this nationalistic communitarianism is likely to be most problematic to those whose religions or cultures do not easily fit within national norms. Here we need especially to note the implications of the suggestion that jobs and opportunities for further education are to be made contingent on educational attainment. Such an alignment of incentives now threatens educational and economic exclusion to those who fail to pursue a curriculum that they might see as an instrument of cultural imperialism. To put the matter differently, this nationalistic communitarianism has sought to confront a stronger interest in liberty with stronger claims about the public interest in education. Regardless of whether these claims can be sustained, there is no doubt that the kind of liberty interest eroded is more fundamental and that the level of coercion involved in any articulation between educational attainment and further education, jobs or income is correspondingly higher.

The picture of people as culturally embedded that has begun to emerge in this description is suggestive of a another approach to student engagement than the one often assumed by educational reformers. This approach emphasizes paying more attention to student alienation and the social and cultural conditions of student engagement. It is widely claimed that students are disengaged from schooling (Goodlad, 1984; Sebring, 1996; Steinberg, 1996; and Toch, 1991) They do not experience the school's goals or its culture as their own. The idea that people are culturally embedded permits an account of alienation predicated on the assumption that people are socially formed (Taylor, 1992). Every human being is born into a cultural context. This culture provides both a sense of identity and a sense of orientation to life. It may be associated with something like a religion or a moral tradition, but it need not be. To say that people are socially formed is to say that their enculturation will have given them a sense of who they are and who are theirs, and it will have initiated them into an orientation to what is worthwhile in life (Taylor, 1989). People have a strong interest in their culture not only in that culture is a background condition of the choices available to people, but also in that culture forms their capacity to choose (Kymlicka, 1995).

If so, then educators who wish to engage students need to think less in terms of incentives to get students to comply with adult expectations and need to be more concerned to ask questions as "How can we constitute schools as communities in which students see themselves as participants in a shared effort in the pursuit of worth while activities?" "How can we provide an education that is culturally appropriate and extends and shapes the values that students bring to schools in ways that they come to value what is educationally worth while?" and "How can we build relations of trust between the adults and the students in a school so that students find the fact that adults value something to be a reason why they should?" Viewing students as socially constituted and culturally embedded instead of seeing them as bundles of preferences tends to transform the question of engagement from one of providing incentives to one of building community. However, it is not clear that what is required is a national community. As a remedy for disengagement, educators need to build on the culture that students bring to school with them, not on one reformers think they ought to have. (For a sketch of a communitarian view of student engagement see Strike, 1991a.)

While I have developed the idea that people are socially constituted and have identities and values formed by their cultures out of the discussion of cultural conservativism, the argument that leads to a desire to form a national culture requires two basic premises. The first is that people are culturally embedded, and the second is that it is difficult for human beings to cooperate in civil and political society unless they are also, in some way, a common people. (For discussion see Kymlicka, 1995.) If we reject the second claim while continuing to hold the first, we may come to a view that regards how people have been enculturated as a important factor in how they are educated, but which also is consistent with a more robust pluralism that is more accepting of diverse cultures and less inclined to want to create a common national culture. The picture presented by Rawls in which society is viewed as a "union of social unions" (Rawls, 1971) and in which political society is held together by a shared view of justice, not a broadly based shared culture, is more consistent with this understanding of multiculturalism than is either the economic picture of human beings that sees them as bundles of preferences or the cultural conservatives picture which seeks the ties that bind in a shared national culture.

This discussion suggests that centralized goal formation and systemic reform may involve some unarticulated and inconsistent assumptions about human nature and human motivation which assumptions may point to doubtful views about how schools are to interact with students. The emphasis on incentives seems at least loosely associated with the economist's picture of human beings as bundles of preferences and suggests a picture of engagement that seeks to manipulate student engagement through external rewards. The emphasis on a shared culture sees people as culturally embedded, but proposes a shared national culture that will be experienced by many as a form of cultural imperialism. Standards driven, systemic reform, may lack a view of student engagement that addresses student alienation and which is decently multicultural.

I do not think that the educational agenda represented by what I have called the conservative compromise is coherent enough to be successfully pursued over a long period. (But consistency does not always rule.) The first and obvious reason is the tension between neo-classical economics and cultural conservatism. And in the American context it is unlikely that there can be agreement on the culture to be conserved. Alan Bloom and Pat Robertson may wish to conserve a culture, but surely they lack a shared idea of what it is. Also, so far as educational standards are concerned, the agenda of cultural conservatives is likely to be inconsistent with the professional cultures of the various scholars and educators who are likely to execute the task of developing educational goals and who may well produces standards unacceptable to cultural conservatives. The considerable hostility to the history goals seems a forceful illustration. Moreover, one important component of the current Republican coalition is very much of two minds about all of this. On one hand the desire to promote school prayer, teach creation science and abstinence, to elect conservative Christians to school boards, and, broadly, to reclaim public schools for a Christian America might be viewed as consistent with the agenda of cultural conservatism. On the other hand, the desire of the religious right to opt out of the public schools at public expense via vouchers is also strong. Many among them have become adept at making liberty arguments in favor of such arrangements (Baer, 1993). The similarity of these arguments to those made by some multiculturalists for schools emphasizing their cultures is evident (Kirp, 1991). I doubt that religious conservatives can make such arguments consistently without legitimating a demand for educational pluralism more broadly. (But consistency does not always rule.) Finally, while conservatives may find something to like in centralized goal formation, they are unlikely to transfer this affection to systemic reform. That they will see as so much big government and as the continued rule of public education by intellectual elites.

Summary: This discussion of modern conservatism has provided another window on the liberties and public interests that might be at stake in standards driven systemic reform. We can represent the liberty interests as our interest in freedom of association, freedom of religion, and the freedom to form our own conception of the good and to pursue it with the like minded. Arrayed against these interests is a kind of communitarian nationalism that sees a widely shared American culture as a requirement of democracy and of social stability. Not only have we altered the conception of the liberty interests and the public interests at stake, we have also altered the picture of human nature. We have moved from a picture of human beings as bundles of preferences to a picture of people as culturally embedded.

I have also suggested that centralized goal formation might be smiled upon by conservatives because it promises to provide markets without pluralism. It thus suggests an educational program that can fuse the interests of two inconsistent forms of conservative ideology. However, I have also argued that the conservative compromise is unstable.

This analysis suggests a key dilemma that standards driven systemic reform must resolve. On one hand, its plausibility against market views of schooling depends on its ability to assert and pursue goals other than efficiency in human capital production and equal opportunity. Goals such as tolerance and democratic citizenship that are plausibly associated with public education seem required. However, if it is to avoid the charge of cultural imperialism, it also needs to take care in linking the agenda of reform to aspirations for a national culture. Such a view is inconsistent with reasonable pluralism and is likely to prove unstable.

Conclusions

The main arguments of this paper can be summed up in two ideas.

> 1. Standards driven, systemic reform raises a number of serious questions about liberty. It may be inconsistent with some liberty interests of students. It is likely to pose serious questions about academic freedom and about pluralism. These concerns should make us cautious about systemic reform and should motivate us to a broader discussion of its assumptions and consequences and to a less ambitious view of its implementation.
>
> 2. The liberty interests at stake in standards driven systemic reform can be represented in at least two ways. Within the framework of neo-classical economics, the essential tension is between a view that emphasizes consumer choice because it is efficient and because it makes schools attentive to the preferences that people actually have and a view in which the state, at a central level, determines the ends to be sought and seeks efficiency through improved public administration. The second framework see the tension as one between freedom of association and pluralism and a kind of nationalistic communitarianism. I have argued that the first framework does not give an adequate picture of

> what is at stake and that a defense of any view of reform, but especially systemic reform, that emphasizes pursuing efficiency in the public sector needs to assert a stronger set of values than those associated with human capital formation. I have also suggested that the conservative view of reform that links neo-classical economics with a kind of nationalist communitarianism is potentially oppressive and unstable.

Has the argument provided a general case against standards driven, systemic reform. Surely not. But these arguments do suggest a framework in terms of which a view of systemic reform needs to be articulated. Systemic reform needs to accept two basic premises. First, it needs to recognize that arguments capable of defending a public sector emphasis in reform need to assert not just human capital formation as the end of education, but also political values such as democratic citizenship or tolerance. Second, it needs to avoid articulating these political goals within a framework of nationalistic communitarianism. What I believe is called for is a conception of reform that takes seriously the commitment to what I earlier called "centralized localism" and that seeks to broaden it to include pluralism as well as localism. The key to centralized localism is to recognize where it makes sense to weaken demands for centralized goal formation and where it does not. If we are to have an educational system we need to emphasize clearly that and how its is to be a flexible system capable of local variation and respectful of diversity.

How? The arguments of this paper suggest a few directions that should be considered as well as questions to be asked.

First, we need to ask how we can encourage a climate of excellence by means other than centralized goal formation and high stakes tests. The idea is hardly absurd. It is what good universities and public schools have often found ways to do. Developing standards might be part of such an effort so long as we develop alternative visions of voluntary standards (McLaughlin and Shepard, 1995) and are restrained in how we embed standards in a network of requirements and obligations. We need to think more in terms of providing models of excellence and less in terms of designing a system.

Second, we need a picture of reform more consistent with pluralism. Rawls's picture of reasonable pluralism has more to offer us than the picture of a national culture envisioned by cultural conservatives. We need to be satisfied with people who pursue excellence in ways consistent with their own cultures or religions so long as legitimate public interests are also satisfied. We need to be less concerned with what people do in schools and more concerned that they do it well. And we need to tailor accountability requirements so that they insist only on what is clearly justified by the public interest and so that they do not preclude reasonable pluralism. A key feature of such an approach may be to develop some assessment and accountability measures that emphasize accountability to parents and the local community as well as or instead of to the state. This would be important in those areas where pluralism is most important. Accountability to the state might emphasize seeking out areas of intersection between different views of education and asking for high performance on essential areas of knowledge.

A flexible approach to accountability is crucial. Two slogans might be useful. One is that *the state should seek to set a high, but narrow bar*. Here the height of the bar is the level of achievement expected. The width of the bar is the range of what schools or students are responsible to the state for. Setting a high, but narrow bar means that governments should expect schools and students to meet high standards, but only for those topics that are clearly linked to the public interest. (Literacy is an example. It is central both to human capital development and citizenship, and it will be a part of any plausible view of education.) In other areas the state might seek ways to inspire excellence, but should avoid accountability mechanisms that detail that in which it consists. Detailed history standards are a mistake when they are so detailed so as to preclude appropriate local variations. A second slogan is that *in areas where diversity is important but where the public interest is also involved, states should promote standards that are thick, but vague*. (History, again, is an example.) Thick vague standards would seek to characterize and illustrate an approach or approaches to some subject matter, but would also do so in a way that permitted local diversity in implementation. Assessments, if they are deemed necessary, should be designed so as to not undercut such local diversity in implementation. Standardized tests should probably be avoided. A decent respect for pluralism may mean restraining the need to hold schools accountable to state government for everything they do.

Third, states need to provide exemptions to standards when schools are able to show that they have good reasons to pursue alternative approaches. These might range from course specific exemptions to charter schools.

Fourth, while it is a central responsibility of schools to produce good citizens, we need a conception of citizenship that emphasizes justice, mutual respect, reasonableness and reciprocity, but which carefully avoids explicating the requirements of citizenship in terms of a thick, shared national culture.

Finally, we need a better approach to the disengagement of students than one that emphasize external incentives and high stakes tests. Setting high standards will do us little good if students have no concern with meeting them (Toch, 1991). But asking how we can get students to be motivated to pursue something educationally worthwhile because the education they are asked to pursue is internal to a worthwhile conception of life that they share is a better question than asking how we can generate external incentives to get students to learn what we believe they should learn.

The key to getting this right is to have a vigorous public discussion about the public purposes of education and how those purposes might be accomplished in ways that respect pluralism and localism. A crucial part of such a discussion is how we may balance the public and the private purposes of education and freedom of association and geography as ways of constituting school communities. Another is how we can provide for vigorous local participation in the full range of educational decisions.

I do not intend this paper to be a blanket endorsement of educational decentralization or localism. Some local agencies will be incompetent or even corrupt. Some reforms, especially fiscal reform, require centralization. Moreover, local jurisdictions can be oppressive and may need to be restrained by central government agencies. Slavery, after all was once a local option. And it is, after all, federal courts that protect our civil liberties often against local jurisdictions. Moreover, what I have argued here seems to me to be perfectly consistent with a vigorous state role in ensuring that all schools meet reasonable minimum standards and with accountability to excellence in subject areas narrowly construed to emphasize the public interest. Nevertheless, the messiness of localism may, in some cases, protect our liberty and our pluralism.

We need a broader debate -- one in which the influence of systems theory is tempered by a deeper concern for a wider range of educational and political considerations. We do not want to wake up some day to find that standards driven, systemic reform has given us an efficient educational system inconsistent with the liberties of a free people.

Post Script: Confessions of a Liberal Communitarian

I have tried to construct the arguments in this paper so as to appeal to widely shared political values central to all liberal democratic societies. However, my concerns about centralized goal formation and systemic reform flow from a more distinctive set of political and educational commitments as well. Here I want to outline these. What follows may help the reader understand why I care about these issues and to avoid projecting an erroneous set of commitments on these arguments.

Communitarians (of my sort) claim that much of human life is and ought to be lived in particularistic communities and is and ought to be conceptualized within particularistic moral traditions. By "particularistic," I mean that these communities and their values and traditions are not and need not be shared with everyone in the larger society. Examples of such communities are (or might be) churches, voluntary fire associations, extended families, small towns, ethnicity based institutions, choirs, service agencies, neighborhood pubs, unions, and work places. Examples of particularistic traditions are religions, cultures, philosophies, and conceptions of human flourishing.

Particularistic institutions are the places in which we form valued human connections, where we teach, refine and pursue shared values and commitments, and where we develop the capacities that are important to the meaningfulness of human relationships and to excellence and competence in shared pursuits. They are important sources of citizenship. Particularistic moral traditions are repositories of intellectual and cultural resources which enable us to articulate, reflect on, and transform views of a good life. These resources often help us to understand what we are trying to accomplish in our particularistic associations. Sometimes, as in the case of religions, a moral tradition may be constitutive of a community. More often, as in the case of choirs or pubs, this is not the case. Nevertheless, moral traditions are important to understanding what we are doing in these communities. Communities in turn are important because they form human connections and socialize people into moral traditions and valued practices.

I think of myself as a communitarian because I believe that these particularistic institutions and traditions are essential to well lived human lives. My kind of communitarianism is in part a view of what some of the goods of human life are (they are human relationships and the common pursuit of shared goods), but it is more a view of how people ought to construct a view of a good life. Good lives are not best constructed by a process in which untutored children shop through a cultural cafeteria of goods filling up their personal shopping basket with assorted values. (For an example of such a view example see Ackerman, 1980.) Building a picture of the good life is more a matter of initiation into practices human beings have found valuable, of applying the resources of some moral tradition(s) to the choices life presents, and of subjecting the result to criticism and debate. (See MacIntyre, 1988; Strike, 1991a; and Strike, 1994.) Moral traditions are resources for the examined life. Apart from them children are not so much free as they are victims of untutored desire.

I am a liberal because I believe that participation in particularistic communities should not be coerced by the state and that a society needs to be bound together by a shared political culture the center of which is a shared view of justice. In order to be widely shared, this view of justice needs to be constructed so as to not privilege any one of the various competing moral traditions or views of a good life so long as they are consistent with justice. Liberal justice is focused on a fair distribution of social resources so that people have an equal opportunity to pursue a self chosen conception of their lives. While a liberal view of justice may resist views of the good that are unjust, it must otherwise seek a decent and principled neutrality between them. I view both freedom of religion and multiculturalism as expressions of this notion of neutrality. Unlike some liberals, I also believe that we are unlikely to succeed in helping our children to achieve a sense of justice if we do not first succeed in initiating them into some worthy, but particularistic culture or moral tradition.

One understanding of democracy sees the primary function of democratic institutions as vectoring interests. People have various interests which they express through their votes. In order to get elected, candidates must create a package of policies combining many of the interests of voters so as to create a majority among the electorate. Similarly in order to get legislation passed, legislators must integrate interests in such a way that there is a majority in favor of some policy. When this process works well, policies result that give as many people as much of what they want as is practicable. A good democracy is an effective bundler of preferences.

A deficiency of this view is that it does not legitimate a means whereby the interests or the views of a good life that underlie them, can be made objects of criticism or discussion. My preference for a more deliberative view of democracy is essentially a preference for an understanding of democracy in which criticism of interests and views of the good life is a part of the political process. (For discussion see Benhabib, 1996.) The purposes of such criticism are to subject views of the good that are unjust to effective criticism, and, second, to promote more effective deliberation on the nature of good lives.

These comments suggest several tasks that a good education ought to accomplish.

1. Every child needs to be initiated into some moral tradition(s) or cultural outlook(s) that provide cognitive resources for reflecting on a good life.
2. Every child needs to develop a sense of justice, to be initiated into the civic culture, and to develop the capacity for democratic deliberation.
3. Every child needs to be acquainted with and to engage in dialogue with other traditions and cultures. The purposes of this are the development of tolerance and the opportunity to assess one's own culture or moral tradition.

These views might be institutionalized in various ways, but they require three things of an educational system. First, particularistic traditions need to be more important than they now are. This requirement might be met by a higher level of care on the part of schools not to compete with a child's church or religion. It might be accomplished, in part, through a diversity of extra curricular activities of the sort legitimated by the Equal Access Amendments and *Westside v Mergers*. Such practices bring freedom of association within the common school. I also suspect that for minority cultures it will require a higher degree of control over the education of their own children. The opportunity costs of cultural transmission are higher for minorities (Kymlicka, 1989). In some cases, it may require opportunity for people to fully operate their own schools so long as the public interest is also served. I do not know the mix of public and private institutions that might be optimal here. I would assume that this would much depend on the mix of cultures, races, and religions in an area. There are both public and private purposes of education, but we should not uncritically assume that private purposes can be served only by private institutions or public purposes by public institutions.

Second, schooling needs to be infused with more local and discursive forums. Such forums are educative because they promote democratic participation and because (1) participation is essential to developing democratic character, (2) local face to face dialogue is the best way to meet those who are different from ourselves, to learn to respect them as equals, and to consider their views in relation to our own, (3) participation is community constituting, and (4) participation promotes a sense of ownership.

Third, schools need to provide real opportunities for students to discuss and argue about different views of a good life.

Forth, we need to think of engaging students in their own education in ways which emphasize the initiation of students into moral traditions which will enable them to see the value of the education adults wish to provide.

To see why I see these aspirations are in tension with comprehensive centralized goal formation consider two different pictures of national standards.

One picture sees national standards as an expression of a kind of nationalist communitarianism, of a view that believes that we need more of a shared culture than is likely to be provided through the attempt to create a shared political culture built around a shared view of justice. Alternatively, we might claim that national standards assert only content that is objectively true or widely held and that is, therefore, neutral to our diverse cultures and moral and religious traditions. Math, it seems oft said, is everywhere the same.

Here there is a dilemma. On the first view, national standards are too thick (on liberal standards) and as a consequence they are oppressive to minority cultures or moral traditions. On the second, they are too thin. They do not accomplish the initiation of students into the moral tradition(s) of some particularistic moral community(s) or put these traditions on the table for discussion and debate. If they dominate the time and resources for education, they may prevent them from being accomplished elsewhere.

Unless national standards can be made consistent with a fairly robust form of what I have called "centralized localism" they will be either too thick or too thin. Like our civil religion - a faith too thin to satisfy the real purposes of a serious theological tradition, but thick enough to offend the religious sensitivities of many - they may be simultaneously too thick and too thin. In a deeply pluralistic society such as ours a centralized curriculum will have considerable difficulty evading this dilemma.

Localism is important in other ways that standards and systemic reform may inhibit. The standards movement takes discussions about what is educationally worthwhile - about goals and curriculum content - and locates these discussion in national or state forums. What is left to locales is implementation. Because it constrains local participation, the educative and the community constituting functions of local discussions is diminished.

Also the standards movement and systemic reform abet the tendency to conceptualize schooling as a government service to be efficiently provided to clients. Parents and students are thus not likely to be seen as community members, citizens, participants, or owners of the education to be provided. While inefficiency is no virtue and efficiency no vice, we should consider whether the framework in which questions about efficiency are asked poses the issues in a wrong or misleading way. Possibly in an educational system in which there is a stronger sense of community and ownership, efficiency will be easier to achieve. Perhaps we need to focus more on what makes schools good communities and less on what makes them efficient organizations.

Finally, the standards movement and systemic reform view students as consumers of a service and as resources for national productivity wars. It sees them as having interests in further education, jobs and income. Thus it seeks to motivate them to serve the collective interest in productivity and their own interest in employment and income by linking these interests to the mastery of a centrally defined curriculum. Nothing in these proposals seeks either to transform student interests or to reduce their level of alienation and disengagement. Instead, it generates incentives to comply with the system's expectations. A more communitarian approach would seek the causes of disengagement and seek to motivate students by helping them to become participants in shared purposes.

It is noteworthy that the Chicago school reform initiative, one that self consciously emphasizes democratic localism, is one of the few visible reform efforts to link student engagement to any notion of community (Sebring, 1996). In contrast one of the more visible recent discussion of student disengagement, Steinberg's (1966), is largely silent on the communal aspects of schooling preferring the use of high stakes tests as a principle device of reengagement. This is particularly perplexing in that one of the main causes of student disengagement is held to be the peer culture, an analysis that seems to invite a more communitarian solution.

I regard the tendency to conceptualize the debate about reform as between those who emphasize markets or quasi-markets and those who want to make the public sector more important as an unhappy construction of our choices. The primary choice is not between market and state. It is how we can have schools that are educative communities that serve both the public and private functions that schools need to serve.

Perhaps the current system of public schools does not serve these aspirations well and systemic reform is unlikely to make things worse. Moreover, my aspirations for schooling are not easily achieved in a modern nations where so many forces undermine community. Perhaps they are beset with what one author calls terminal wistfulness (Stout, 1988). But we will never know unless we try, and we will never try if we do not give the matter some thought. My deepest concern with standards and systemic reform is that they focus attention on the wrong questions and that they perpetuate our inability to ask the right ones.

References

Ackerman, B. (1980). *Social justice in the liberal state*. New Haven, CT: Yale University Press.

Adler, M. J. (1982). *The Paideia proposal*. New York: Macmillan.

A New Compact for Learning. (1994). Albany: The University of the State of New York The State Education Department.

Arons, S. (1994). The Threat to Freedom in Goals 2000. In N. Cobb

(Ed.), *The Future of Education* (pp. 55-60). New York: The College Board.

Baer, R. A., Jr. (1993). "Strict Neutrality" and Our Monopoly System. In J. W. Skillen (Ed.), *The School Choice Controversy* (pp. 15-34). Washington, D.C.: Baker Books.

Benhabib, S., Ed. (1996). *Democracy and Differences: Contesting the Boundaries of the Political*. Princeton, NJ: Princeton University Press.

Bennett, W. J. (1984). *To Reclaim a Legacy: A Report on the Humanities in Higher Education*. Washington, D.C.: National Endowment for the Humanities.

Berliner, D. C. (1993). Educational Reform in an Era of Disinformation. *Educational Policy Analysis Archives*, 1(February 2) (Online: http://olam.ed.asu.edu/epaa/).

Bishop, J. (1989). Why the Apathy in American High Schools? *Educational Researcher*, 18 (January-February 1989), 6-10, 42.

Board of Regents v Bakke, 98 S. Ct. 2733 (1978).

Callan, E. (1996). Political Liberalism and Political Education. *Review of Politics*, 58 (1), 1-33.

City on A Hill Charter School, http://www.doe.mass.edu/cs.www/cs.cityonahill.html.

Clune, W. (1993). Systemic Educational Policy: A Conceptual Framework. In S. H. Fuhrman (Ed.), *Designing Coherent Educational Policy* (pp. 125-140). San Francisco: Jossey -Bass.

Coleman, J. S. (1974). *Youth: Transition to adulthood*. Chicago, IL: University of Chicago Press.

Finn, C. E., & Walberg, H. J. (Eds.). (1994). *Radical Education Reforms*. Berkeley, CA: McCutchan.

Finn, C. E. Jr. (1996, February 14). On Governors and Ostriches. *Education Week*, pp. 44, 34.

Fonte, J. (1995, June 9). The Naive Romanticism of the History Standards. *The Chronicle of Higher Education*, p. 48.

Friedman, M. (1962). *Capitalism and Freedom*. Chicago: University of Chicago Press.

Fuhrman, S. H. (1993). The Politics of Coherence. In S. H. Fuhrman (Ed.), *Designing Coherent Educational Policy* (pp. 1-34). San Francisco: Jossey-Bass.

Gagnon, P. (1995, December). What Should Children Learn. *The Atlantic Monthly*, 65-87.

Gardner, H. (1982). *Art, Mind and Brain*. New York: Basic Books.

Gardner, H. (1983). *Frames of Mind*. New York: Basic Books.

Goals 2000: Educate America Act.(1993).

Goodlad, J. I. (1984). *A place called school: Prospects for the future*. New York: McGraw-Hill.

Gutmann, A. (1995). Civic Education and Social Diversity. *Ethics*, 105 (13), 557-579.

Hanushek, E. (1995). Outcomes, Costs, and Incentives in Education. In N. R. Council (Ed.), *Improving the Performance of the Educational System* . Washington, D.C.: National Academy Press.

Hirsch, J., E. D. (1987). *Cultural literacy: What every literate American needs to know*. Boston, MA: Houghton Mifflin.

Kirp, D. (1991). Textbooks and Tribalism in California. *The Public Interest*, 104, 20-36.

Kuhn, T. (1970). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Kymlicka, W. (1989). *Liberalism, community, and culture*. Oxford, England: Clarendon Press.

Kymlicka, W. (1995). *Multicultural Citizenship: A Liberal Theory of Minority Rights*. Oxford: Clarendon Press. Clarendon Press.

Lakatos, I. (1970). Falsification and the Methodology of Scientific Research Programs. In I.Lakatos and A. Musgrave (Eds.), *Criticism and the Growth of Knowledge*. (pp. 91-196), London: Cambridge University Press.

Lee v. Weissman, 505 U.S. 577 (1992)

Levittown UFSD v. Nyquist [57 NY2d 27] (1982)

Macedo, S. (1995). Liberal Civic Education and Religious Fundamentalism: The Case of God v. John Rawls. *Ethics*, 105(3), 468-497.

MacIntyre, A. (1988). *After virtue*. Notre Dame, IN: University of Notre Dame Press.

Malen, B., & Ogawa, R. (1990). Community Involvement: Parents, Teachers, and Administrators Working Together. In S. Bacharach (Ed.), *Educational Reform: Making Sense of It All* (pp. 103-120). Boston: Allyn and Bacon.

Manzo, K. K. (1997). In Mass., Silber Goes to Battle Over History Framework Draft. *Education Week*. Washington: pp. 1, 19.

Mathematics, N. C. o. T. o. (1989). *Curriculum and Evaluation Standards for School Mathematics*. Reston, Va.

McCarthy, T. (1994). Kantian Constructivism and Reconstructivism: Rawls and Habermas in Dialogue. *Ethics*, 105 (1), 44-63.

McLaughlin, M. W., & Shepard, L. A. (1995). Improving Education Through Standards-Based Reform . Stanford, CA: National Academy of Education.

Mill, J. S. (1956). *On Liberty*. New York: The Bobbs-Merrill Company, Inc.

Milliken v. Bradley, 418 U.S. 717. (1974).

Monk, D., H. (1990). *Educational Finance: An Economic Approach*. New York: McGraw Hill.

National Commission on Excellence in Education. (1983). *A Nation at Risk: The Imperative For Educational Reform*. Washington, D.C.: The Commission.

New York State Education Department (1995). Curriculum, Instruction, and Assessment: Preliminary Draft Framework for Social Studies.

Noddings, N. (1992). *The Challenge to Care in Schools: An Alternative Approach to Education*. New York: Teachers College Press.

O'Day, J. A., & Smith, M. S. (1993). Systemic Reform and Educational Opportunity. In S. H. Fuhrman. (Ed.), *Designing Coherent Educational Policy* (pp.250-312). San Francisco: Jossey-Bass.

Pierce v. Society of Sisters of the Holy Name of Jesus and Mary, 268 U.S. 510, (1925).

Okin, S. M. (1994). Political Liberalism, Justice and Gender. *Ethics*, 105 (1), 23-43.

Paris, D. C. (1995). *Ideology and Educational Reform*. San Francisco: Westview Press.

Rawls, J. (1971). *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Rawls, J. (1993). *Political Liberalism*. New York: Columbia University Press.

Reich, R. B. (1990). Preparing Students for Tomorrow's Economic World. In S. B. Bacharach (Ed.), *Educational Reform: Making Sense of It All* (pp. 194-233). Boston: Allyn and Bacon.

San Antonio Independent School District v. Rodriguez, 411 U.S. 1 (1973).

Saxon, J. (1995, December 6). Arrogance Run Amok. *Education Week*, (pp. 8,9).

Saxon, J. (1996, February 7, 1996). The NCTM Standards are a Very Bad Joke. *Education Week*, (pp. 27,28).

Scheffler, S. (1994). The Appeal of Political Liberalism. *Ethics*, 105 (1), 4-22.

Schlesinger, A. M. J. (1992). *The Disuniting of America*. New York: W.W. Norton & Company.

Sebring, P. B., A. S. Bryk et al. (1996). Charting Reform in Chicago: The Students Speak. Chicago, Illinois: Consortium on Chicago School Reform.

Shanker, A. (1994). National Standards. In C. E. Finn, Jr. & H. J. Walberg (Eds.), *Radical Education Reform* (pp. 3-20). Berkeley, California: McCutchan Publishing Company.

Smith, M. S., Scoll, B. W., & Link, J. (1996). Research-Based Reform: The Clinton Administration's Agenda. In N. R. Council (Ed.), Improving the Performance of America's Schools (pp. 9-27). Washington, D.C.: National Academy Press. Administration's Agenda. In N. R. Council (Ed.), Improving the Performance of America's Schools (pp. 9-27). Washington, D.C.: National Academy Press.

Steinberg, L. (1966). *Beyond the Classroom*. New York: Simon and Schuster.

Strike (1979). An Epistemology of Practical Research. *Educational Researcher* 8(1): 10-16.

Strike, K. (1982). *Educational policy and the just society*. Urbana: University of Illinois Press.

Strike, K. A. (1984). Fiscal Justice and Judicial Sovereignty: Plotting the Logic of a Slippery Slope. *Educational Theory*, 34(1), 5-21.

Strike, K. A. (1990). Is teaching a profession: How would we know?. *Journal of Personnel Evaluation in Education*, 4, 91-117.

Strike, K. A. (1991a). Humanizing education: Subjective and objective aspects. *Studies in Philosophy and Education* 11(1): 17-30.

Strike, K. A. (1991b). Liberal Justice: Aspirations and Limitations. In D. Verstagen & J. Ward (Eds.), *Spheres of Justice in Education: The 1990 American Education Finance Association Yearbook*. New York: Harper Business.

Strike, K. A. (1991c). The Moral Role of Education in a Liberal democratic Society. In G. Grant (Ed.),*Review of Research in Education* (Vol. 17, pp. 413-483). Washington, D.C.: American Educational Research Association. (Ed.), Review of Research in Education (Vol. 17, pp. 413-483). Washington, D.C.: American Educational Research Association.

Strike, K. A. (1993). Professionalism, Democracy, And Discursive Communities: Normative Reflections On Restructuring. *American Educational Research Journal*, 30 (2), 255-275.

Strike, K. A. (1994). On the construction of public speech: Pluralism and public reason *Educational Theory* 44(1): 1-25.

Stout, J. (1988). *Ethics after Babel*. Boston, MA: Beacon Press.

Taylor, C. (1989). *Sources Of The Self: The Making Of The Modern Identity*. Cambridge, Mass.: Harvard University Press.

Taylor, C. (1991). *The Ethics of Authenticity*. Cambridge, Mass.: Harvard University Press.

Taylor, C. (1992). *Multiculturalism and "the politics of recognition"*. Princeton, NJ: Princeton University Press.

Toch, T. (1991). *In The Name Of Excellence*. New York: Oxford University Press.

Tyack, D., & Cuban, L. (1995). *Tinkering Toward Utopia*. Cambridge. Massachusetts: Harvard University Press.

Verstagen, D. (Ed.). (1991). *Spheres of Justice in Education: The 1990 American Education Finance Association Yearbook*. New York: Harper Business.

Vitz, P. (1986). *Censorship: Evidence of Bias in Our Children's Textbooks*. Ann Arbor, Michigan: Servant Books.

Wise, A. (1990). Overregulating Our Schools. In S. Bacharach (Ed.), *Educational Reform* (pp. 400-409). Boston: Allyn Bacon.

Young, I. M. (1990). *Justice And The Politics of Difference*. Princeton, New Jersey: Princeton University Press.

About the Author

Kenneth A. Strike

Kenneth A. Strike is Professor of Philosophy of Education at Cornell University where he has taught since 1971. He has a B.A. from Wheaton College and M.A. and Ph. D. degrees from Northwestern University. He has been a distinguished visiting professor at the University of Alberta, is a past president of the Philosophy of Education Society and is a member of the National Academy of Education. His principal interests are professional ethics and political philosophy as they apply to matters of educational practice and policy. He is the author of several books and over a hundred articles. Recent works include *The Ethics of Teaching* (with Jonas Soltis), *The Ethics of School Administration* (with Jonas Soltis and Emil Haller), *Liberal Justice and the Marxist Critique of Schooling, Ethics and College Student Life* (with Pamela Moss), "Professionalism, democracy and discursive communities: Normative reflections on restructuring" in the *American Educational Research Journal*, "The moral role of schooling in a liberal democratic society" in the *Review of Research in Education*, and "The Moral Role of Educators" in *Handbook of Research on Teacher Education*. He has had a durable interest in questions of religious liberty and schooling and in using liberal political philosophy as a lens for conceptualizing educational issues. His current work emphasizes the exploration of the normative aspects of school reform. Also Professor Strike is an enthusiastic canoeist, and, while he duly notes the academic irrelevance of this fact, he mentions it because it speaks to his good character, aesthetic sensitivity, sound body, and (possibly) sense of humor.

Email: kas14@cornell.edu

Kenneth Strike's Homepage.

Copyright 1997 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://olam.ed.asu.edu/epaa

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey@olam.ed.asu.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Marshall University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Rocky Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
Arizona State University

Dennis Sayers
University of California at Davis

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina--Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher Education

Mary P. McKeown
Arizona Board of Regents

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois--UC

Robert T. Stout
Arizona State University

---

archives | abstracts | editors | board | submit | comment | subscribe | search

EPAA
• enter the archives • browse the abstracts • the editors • the edit board
• submit article • submit commentary • search • subscribe
volume: • 1 • 2 • 3 • 4 • 5 • 6 • 7

This article has been retrieved 1927 times since April 30, 1997.

Education Policy Analysis Archives

Volume 5 Number 12 April 30, 1997 ISSN 1068-2341

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass Glass@ASU.EDU.

College of Education

Arizona State University,Tempe AZ 85287-2411

Copyright 1997, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

Studying the Rural in Education: Nation-Building, "Globalization," and School Improvement

Craig B. Howley
Appalachia Educational Laboratory

Abstract This essay maintains that nation-building, partly through systems of schooling, has served rather more to debase than improve the rural circumstance. It suggests that a different logic of improvement is needed in rural education, but refrains from prescriptions. Instead, it focuses its attention on the sort of questions that researchers (and school improvers, for that matter) might ask to discover or invent that logic variously. It draws a distinction between cosmopolitan and local interests and provides examples of issues that exhibit the distinction. Finally, it suggests and provides hypertext links to sources in sociology, literature, philosophy, and education that might help educational researchers (and anyone else with an interest in "the rural") ground their studies and their actions in issues that honor rural interests. I remind readers that the very word "essay" means "tentative."

Since the 1983 debut of *A Nation at Risk*, good critiques have unmasked the domestic implications of the idea that the nation can be preserved through educational reform. Until recently, however, few have doubted the sustainability of schooling as an institution for nation building, and fewer still have framed such doubts in light of the rural circumstance. The rural experience, however, offers scope for doubt. This essay attempts to raise those doubts from that perspective.

Too often, researchers overlook what is most particularly rural as a fit object of inquiry in educational research. Instead, objects of national--or perhaps, cosmopolitan-- practice absorb their attention and thereby obscure rural issues and dilemmas. Public school curricula and practices do look remarkably similar worldwide (e.g., Benavot, Cha, Kamens, Meyer, & Wong, 1991), and this fact ought to disturb educators more than it apparently does. Educators, after all, have a vested interest in the institutions of cultural transmission. When practices are so similar to one another worldwide, we ought worry about the health of global culture overall. Part of what this essay intends is to pull back the veil of national, or cosmopolitan, concern in order to expose contemporary rural issues not only worthy of attention in their own right, but worthy of particular attention in an ethos more evidently concerned with globalization than with its related need for ruralization (cf. Orr, 1995).

Improvement in rural education requires a logic quite different from the one that still prevails. A changed perspective on the purposes of schooling is needed if educational researchers would help develop institutions that actually benefit rural communities. I will argue here that a worldwide decline in the importance of national sovereignty has made the historic mission of school improvement--conceived as nation-building--counterproductive in rural places. Although few educators have properly understood the relevance of this change to the system of American schooling, there are good reasons why the change is salient to rural communities--and to rural practitioners and researchers. This essay aims to help educational researchers grasp the implications of this change so as to consider issues that are perhaps more germane to rural school improvement efforts.

Rural School Improvement is Nobody's Fault

In the best of all possible worlds, the interaction between disciplined inquiry and the arduous labor of keeping school would naturally, and easily, entail the improvement of schooling. But we no longer trust the naturalness or ease of such an interaction; we are less optimistic about the whole enterprise of schooling, including research about schooling, than formerly. From 1910 to 1965, education was part of the march of progress toward an inevitably better future--a progressive, postwar, and increasingly post-rural future. Although many people behaved as if research had been a part of this improvement, it would probably be fairer to say that ideology and technology were stronger influences. Lately we seem to have discovered that research occupies a seat less exalted than we previously thought. Instead of technology being the handmaiden of research (science), we are more and more frequently acknowledging that research is the handmaiden of technology.

Research cannot, in any case, reliably cause school improvement, because schools improve when those involved jointly *will* it, and this they may do entirely without the benefit of research. Properly speaking, research may *inspire* or *deflate* improvement efforts; that is, depending on its quality, research can either undermine or fortify the will to improve. Of course, research once promised that it would create a *reliable* path to improvement in all realms of life. But it has not been able to live up to this promise in education. As a result, a sort of hysteria for "more practical" educational research has grown up. We forget that the focus of the effort was always on creating reliable processes for improvement. The nature of education, I have argued elsewhere, is such that the goal of fashioning widely applicable and reliable procedures for improvement is unreasonable. This is not the same as saying that improvement is impossible. Clearly, schools do improve and they do decline. What is needed is thoughtful research that attends to the particularities (and not the generalities) of the places of which schools are or ought to be part.

Rural places in the contemporary world may suffer more than other places from the lack of such research and from the misguided effort to build up widely applicable and reliable procedures for school improvement. Too many current research efforts, it seems, undermine improvement in rural education. Too few pay attention to rural circumstances, and too few offer anything to fortify the will of those who would see rural schools improve for the benefit of rural community.

In fact, during the past century and a half, improving rural schools also meant reshaping and redirecting them into a national system--a system of schooling, manufacture, trade, politics, and culture--that has insured, if not required, the depopulation of the countryside. This trend toward rural depopulation and unemployment continues, confounding and subverting rural families and rural communities. The increasing automation of agriculture, mining, and timbering has led to the widely repeated misrepresentation that "rural no longer means agriculture." What people mean by this phrase is that rural *jobholding* no longer means agriculture (or mining or timbering). Technology has effectively severed the natural connections--the economic connections--people once maintained with the land. Fewer rural people--and rural people less and less frequently--are *responsible* for rural land.(1) They are now as "free" as the rest of the nation to wander wherever big capital requires them to go. Schooling is a handmaiden to the system that *is* responsible. And educational research is an (albeit subsidiary) handmaiden to the technologies that enable that national system. Rural school improvement of this sort is nobody's fault; it is endemic and longstanding.

"Today's Culture Sends Strong Messages"

Why has such a subversion of rural places, with its inevitable reflections in practice and in research, continued for so long? CPRE, the Consortium for Policy Research in Education, fingers a culprit sufficiently powerful to explain the problem: "Today's culture sends strong anti-intellectual messages" (CPRE, 1996, p. 7). This, by the way, is what I infer that Wendell Berry (1990) means when he claims that our society is "mind-dominated": We fail to attend to things that might otherwise be obvious to us. The mind turns against itself, and perhaps schooling is the instrument of such a turn of mind. The dilemma of rural education is constituted of this irony.

Not only is it true that anti-intellectual messages "are sent" to us, we are surrounded and relentlessly bombarded by them. Worse still, we repeat them, not only to others, but to ourselves. We live and breathe these messages; the very institutions of our society are anti-intellectual, and our universities are by no means exceptions to this rule (cf. Howley, Howley, & Pendarvis, 1995).

Universities are knowledge factories given to making intellectual work as routine as manufacturing (Anderson, 1993); hence, they are not particularly well-equipped to consider improvements that require breaking rank with established relations of power. Beyond this general shortcoming, universities have generally implemented a cosmopolitan, modernizing mission; and this mission is perhaps pursued with the highest purpose the more the university's territory is understood to be a parochial backwater.

This observation may set some good people's teeth on edge, so let me add that I'm decidedly *not* suggesting that Harvard and Berkeley are the best resources for approaching matters differently in rural education. Their students and professors may empathize and exhibit wonderful insight, but *they are not usually here, or for very long.*

Once more: None of us is singly, or in groups, to blame, but it's as if we've been missing something. What's wrong?

The Default Position of Rural Educational Research

Most educational research, regardless of type of locale, is carried on today as if nation-building were still an operant principle in the world system. This is no wonder, since the most concerted and best funded effort (yielding the best research by technical standards) is sponsored by the federal (national) government.

Several connections can be drawn to prevalent modes of operation under which research into rural education and rural schools is conducted.(2) We should all recall that superbly executed research that fails to ask the right questions will miss the point. C. Wright Mills (1959) took Talcott Parsons sternly to task for this failing ("power" was not a word in Parsons's lexicon).

First, skepticism prevails that rural schools are, or should be, very much different from urban or suburban schools. Underlying this belief, but seldom articulated, is the presumed Platonic ideal form-of-the-school, of which rural, urban, and suburban schools are all equally imperfect images. With us, the common school is not common simply because all neighboring children of whatever class or race have attended it, but it is understood as having the idealized form of the school in common with all other American schools. When Americanization was the essential educational project, such an ideal served a purpose; one could argue--though few do so today--that the ideal was emancipating for the European-Americans who invented it.

Second, beneath this skepticism about rural schools runs a deep disregard of actual rural places, diverse as they are. Disregard of these places entails the invisibility of the peculiarities of rural families, rural ways of living and working, and local rural meanings and knowledge. The anti-intellectual larger culture--which is as evident inside universities as outside them (Anderson, 1993; Barzun, 1959; Hofstadter, 1963; Howley, Pendarvis, & Howley, 1995)--makes it easy for scholars and ordinary citizens alike to avoid appreciating these peculiarities, except for fleeting sentimental purposes. In our disregard for manual labor, an enduring reality in rural America, we have as a people forgotten how much reflection and cleverness good handiwork requires. Relying on machines has helped make the culture stupid. Computers may be carrying this stupidity one step further.

Third, on the terms of nation-building, schools and districts are believed to constitute localities that frustrate reform efforts. CPRE (1996, p. 6), for instance, finds that "state and local agencies are slow to adapt to new policy goals." Presumably these new policy goals are those promulgated at the national level by learned societies and various combinations of national leadership. There is an historical shift of perspective inherent in this view: Where localities in the 19th century could be understood as units that contributed to nation-building, today they are widely understood as impediments to national vigor. In the 19th century, states *joined* the nation. In the 19th century, *Americans* shaped localities and *localities* shaped Americans. Today, local districts, schools, and teachers need to be fixed, systemically, and almost-all-at-once (cf. Sashkin & Egermeier, 1994). In the case of schools in rural places, the very nature of the systemic argument blames the victim while inflicting further punishment.

Fourth, the imperative that research now become exceptionally "useful"--that is, that it technologize the mission of truth-finding and truth telling-- means that studies must be designed to ensure development of educational goods and services that can be marketed as universally effective, or nearly so. In practice, one might argue, this rumor of universal effectiveness promises that most such goods and services will be out of place in most locales. Few researchers believe in the old corporate model of development any longer, but many now hail Total Quality Management as the next wave of development; and "world-class-standards" are the watchword in this new scheme.

The Problem

There are other assumptions in American culture that might be listed as contributing to the difficulty of inquiring about rural schools and communities, their mutual educational dilemmas, and their mutual sustainability. And the conditions under which educational researchers, working in very instrumental ways, might also be mentioned (e.g., timing their labors to election cycles). But the four assumptions given above are enough to illustrate the problem.

If, for instance, the topic is statewide reform, again, the focus of effort is likely to be the special backwardness or challenges of rural places in acceding to the reforms, not the disjunction between local and state priorities. If the topic is student aspirations, the focus is likely to be how rural schools can best "increase" the level of students' aspirations, not the relationship between student commitment to rural life ways and cosmopolitan ways. If the topic is the dropout rate, the focus of effort is most likely to be strategies for retaining or retrieving students rather than the disjunction between rural schools' national purposes and the nature of local rural economies. These are all real examples, disguised to protect us all, for I too have conducted such studies. It is not *easy* to see rural issues, and, as Paul Theobald remarked recently in a private conversation, "The learning curve is very long." This paper and the others in this session are trying to help others shorten that curve. A shorter curve, however, may just be a steeper curve.

The problem concerns motive, commitment, and attention. This essay has so far explained the motive for examining rural dilemmas more closely: the primary purpose of mass education in America is collapsing. The time has come for educational researchers to develop different commitments. This, of course, is already happening, as the pages of *Educational Researcher* attest with almost every issue. Among those different commitments, which really are familiar to many researchers, are local purposes for education--forms of education and their institutional expressions that sustain local communities, especially as thoughtful and active cultures.

Nation-Building

In the 19th century, nation building was the major educational project, and much of it took place in *rural schools*, because over half of all Americans lived in rural locales until 1917 or so. By Americanizing children local rural schools not only converted immigrants into citizens, they ironically also converted themselves as a national system. The ideological roots of the national system are arguably rural (Meyer, Tyack, Nagel, & Gordon, 1979; Theobald, 1995).

Our historical motive for improvement seems to be that we wanted then--and still want-- good Americans to issue forth from schools, like Athena from the head of Zeus, without histories or idiosyncracies. The only adaptation we have made today is that we now want them perfectly prepared for, and accepting of, the emerging postindustrial conditions of employment, which look less promising than the conditions of industrial employment at the height of modernist industrial power in post-war America. The grim outlook we hold on the future, of course, bodes ill for school improvement conducted in the name of building the nation.(3)

This point was recently brought home to me in a book to which I have been alluding (Anderson, 1993). The author, a scholar of political science at the University of Wisconsin, argued forcefully that *improvement* in thought and reality is an aim worth preserving for the university. It strikes me that this argument is needed because we no longer agree about our plans for an idea like "school improvement." Your improvements are my debasements, I might well claim.(4)

The very logic of improvement, it seems, has escaped us. At the end of the industrial age, we doubt the efficacy of reason because *technical rationality* has perhaps unwittingly, but none the less certainly and prominently, helped magnify the evident horrors of the 20th century. Anderson, however, wisely points out that--although progress is not inevitable, nor the workings of power necessarily just--we nonetheless need to believe we *can* make the world a better place and use the powers of the mind to facilitate that end (Anderson, 1993). So reason, it turns out, requires faith to support it.

It is faith we have lost, not reason, and not the need for decent research.

Although our lack of faith undermines the role of K-12 schools in producing good Americans, the purpose has already been rendered specious by the declining importance of nation-states. The educational problem of producing good Americans, aside from the perpetual danger that such a mission is xenophobic, is that *being American* no longer has much to do with shaping the American ethos. The American ethos, according to Christopher Lasch (1994), is being constructed by elites whose primary allegiance is not to American democracy but to their own self-interest in the global marketplace. They wave the flag to pacify, and not to rally, followers.

The work of nation-building, fueled by the 19th century rural experience, is virtually finished. This point requires some interpretation, because it must at first seem short-sighted and improbable, not to mention remote from issues of educational research and practice.

Whereas in the nineteenth century, "we" all wanted to be Americans, we could not then have been so sure what it *meant* to be Americans. To be an American circa 1850, only somewhat less so than in 1750, entailed participation in the project of fashioning (perhaps equally by contribution and by negation) what it meant to be an American. It was difficult work, and we have not found the results very satisfactory; greed, venality, and injustice were as common then as now. As Daniel Kemmis (1990) and others have suggested, the frontier allowed us to hide from ourselves. Rural newcomers, up to and after the closing of the frontier, were acquisitors and not uncommonly thieves. When clever or lucky, they made sure that other newcomers, usually people with fewer initial advantages, kept moving (Theobald, 1995).

The point here is that there is no reason whatever to think that life in America and life in schools would improve if we just returned to the way things were 150 years ago. Nostalgia is an equally bad foundation for educational research and for educational practice.

Nation building, however, is over not merely because America has matured, but because the worldwide project of nation building is ending. The sovereignty of nations is in dispute and under assault, as it must be when national financial resources, which emody national sovereignty, circulate rapidly through international markets ("the global economy"). The political usefulness of the nation-state is being superseded by a new political-economic formation, though little is known about this emergent formation. Sovereignty based on territoriality, however, is less and less functional. It is difficult to determine if, when one "buys American," one really is buying American. Often one is buying "multinational firm," globalized but nonetheless private capital.

Globalized private capital, however, is an unsatisfactory replacement for sovereign territory. The idea of place itself is antithetical--and therefore essential--to the idea of globalization (McMichael, 1996). We need to realize that the emergence of localism in tandem with concern for globalization is not a conundrum or an accident. It represents a new circumstance in the evolution of the world; it also suggests that the focus on global competitiveness constitutes an insufficient preparation for the 21st century.

In short, the way Americans understood ruralness in the 19th century is dramatically different from the way Americans now experience it, and from *the way they must increasingly begin to realize they experience it*. In that phrase ("they must begin to realize they experience it") lies an educational purpose profoundly at odds with 19th and especially 20th century educational purpose. Back then, 'rural' need not have been conscious of itself. It was conscious, rather, of something else, an ideal Americanism in which the rural experience was a passing phase. Today, the rural experience will not continue to exist unless self-conscious and unless it develops its own purposes. But there are many reasons why "the rural" will persist, and more still why it must persist (having to do with what, in Wendell Stegner's fine phrase, people are *for*).

Again, I am not arguing that the rural past was an idyll, nor that rural places are necessary harbors of virtue. Slavery was a rural phenomenon; dispossession of Mexican American colonias was a rural phenomenon. The genocide against American Indians was carried out on the frontier. Although all these injustices contributed to nation building in North America, neither do I argue that nation building is a necessary harbor of evil, nor, further, that emergent political formations may implement justice more properly and more generally (they are more likely to magnify greed and injustice in my view). I do claim that care for particular places is more necessary to justice and decency than ever and that threats to the integrity of local places are multiplying.

Commitments Local and Cosmopolitan

"Local" is a crucial concept to keep in mind when dealing with rural places, as intimated previously. A recent conversation challenged me to say whether or not any useful definition of *rural* were possible.(5) For statistical purposes, it really is easy enough to define rural, but I choose to draw the distinction between *local* and *cosmopolitan*. *Local* can only be described as a set of commitments usefully distinguished from *cosmopolitan* commitments. And because rural places are composed of small communities, with a degree of separateness from one another, in sparsely populated places, they are exemplars of the idea of *local*. And in fact, rustic ways have generally served modernizing society as the perfect antithesis of cosmopolitanism.(6)

Can urban places share such commitments? They can, but, as Jane Jacobs (1984) has pointed out, our cities have been remade by what she calls "the sorters," social engineers and policy makers who have separated industrial zones, residential districts, low-income housing, and commercial districts from one another. The effect of the sorting has been to destroy city neighborhoods--urban localities with parallel commitments to those still prized in rural places--and recreate cities as placeless entities. Cities seem a darker, dystopian, adaptation of the suburban nightmare. Whereas, in the country, people's character is evident in their homes and farms, in the city people try to symbolize their character in modes of self-presentation, particularly the way in which they dress. Changeable fashion substitutes for the enduring substance of place that characterizes rural places and people.

The issue of commitment lies in the realm of ideology: what we believe, why we believe it, whose interests our beliefs advance, and how we reflect those beliefs in thought and action (research and practice). Education for nation building is very different from education that would cultivate particular communities and the ideal of community, on that experiential basis. We are, it is fortunate, much less shy of commitments and values than used to be the case. After all, private businesses, universities, and non-profit corporations are used to developing vision or mission statements founded on what we members assert as beliefs.

This is the place where I am going to get very specific in this essay, because I wish to indicate plainly some of the things that those researchers who might be interested in rural education, but have not yet interpreted its proper interests, could be looking at. These lists are hardly the result of research; they are the result of 10 years of observation of rural literature, however, by someone to whom people have the vexing habit of asking: "What should I study in rural education?" Observe that these are only topics, and that actual studies must consider locally relevant *issues* that emerge from local experiences or that embody local dilemmas. For the sake of contrast I begin with 'cosmopolitan' topics that can be related to rural settings, but which do not share rural *commitments*. (These topics are *not* commended to readers!)

Cosmopolitan commitments.These are cosmopolitan concerns as they apply to rural education: how to...

- increase the level of students' aspirations,
- overcome resistance to consolidation and school closure,
- overcome the disadvantages of students' backgrounds,
- implement state and national reforms,
- offer a broad and deep high school curriculum,
- insulate the school from local politics,
- implement "best practice" (i.e., nationally validated methods and programs), and
- change the local culture.

Rural (local) commitments. The following list is (merely) illustrative of local concerns loosely related to the preceding cosmopolitan concerns, *as they must inevitably be in the era of waning influence among nation-states*, but better capturing issues of the rural circumstance:

- senses of and attachment to rural places,
- the relationship between school and community sustainability,
- proper aims for an education committed to rural community,
- rural pathways to rural adulthoods,
- community engagement in rural schools,
- rural community and educational stewardship,
- curricula to sustain rural places,
- small-scale organization in rural schooling and community, and
- cultivation of appropriate local meanings, knowledge, and commitments.

Most of the usual suspects that researchers and doctoral students in education are advised to consider could be "ruralized" according to the commitments enumerated above. Many educational researchers will first need to separate themselves from their own cosmopolitan training. Several preparations can help.

Preparations

First, there is an excellent rural literature, but *one must read not only in education*, but in rural sociology, rural community development, history, and, especially, literature. Unlike these other disciplines, literature (literary fiction) has typically had a much better grasp of the nation-state *qua* fiction(7) (the Appendix lists books that have influenced me).

Second, and perhaps more critical to one's grounding, is *interest in matters and minds rural*. I mention this preparation second because it ought ideally to go without saying. One nonetheless suspects that some of those who attempt rural studies are more interested in other things (e.g., tenure, the rural venue seeming perhaps less stressful than the national venue). Interest means some combination of experience and understanding that puts one in the middle of the salient real issues. The best experience on which to ground interest (this is a third preparation) may be *living or working in a rural community*. This experience is nonetheless insufficient in itself.

The fourth and perhaps most essential admonition is that those interested in rural educational research should bring some sort of critical framework to the experience. Rural educators who would conduct rural educational research are at risk in this regard, precisely on account of their training, both as educationists and as researchers. The reasons should be clear from the main discussion: Educationists and researchers are cultivated as cosmopolitan nationalists. What am I talking about?

Amitai Etzioni once asserted that the instrumentalism of schooling increases with each higher level, on which basis doctoral programs could be said to be the most instrumental schooling of all. Although this is a strange turn of events for those who are called "doctors of philosophy," very little research is ever guided by philosophy, and the theorizing that does guide it is usually very narrow. This is merely how 'positivism' works. Concerned with, and usually trained to do the thing right, most of us lack not only the personal capacity, but, more importantly, the *institutional* capacity to do the right thing (to borrow a useful phrase from the leadership literature). Our 'training' not only prepares us to ignore local issues, but predisposes us to ignorance.(8)

The Importance of Critique

"Critique" is almost a dangerous word these days; it conjures images of dejected marxists, undisciplined post-positivists, ethereal post-structuralists, Nietzschean postmodern flakes of varied crunchiness, deluded romantic communitarians, and wild-eyed paradigm shifters--many of whom are not only incomprehensible, but who are actually trying to dismiss reality altogether. Academics are held up for special ridicule by those few practitioners and numerous policy makers who know a little of the arcane struggles that beset academe.

The dismissal of critique on this view (i.e., the view that madness lies in the direction of critique) is all rubbish from the perspective of this essay, and it is neither bold nor modern. In fact, responsible and very practical marxist, post-positivist, postmodern, and communitarian critiques are, each of them, very much needed in rural educational research.

Anderson (1993), rather conservative in his outlook, has the role of critique about right: its highest purpose *probably is* in service of action to make the world a better place. This sort of practicality, however, is very different from the national instrumentalism that determines the best-funded research agendas because it allows for realities outside the accepted canon of national security issues. Rural education lies in the shadow of this canon. Critique that aims to make the world a better place might cultivate a supportive, but not directive or determining, expectation for research. Practical critique is needed at present in the construction of properly rural education. Such critique, in fact, has long been needed. Right now, however, the opportunity for it is perhaps more auspicious, since the motive of school improvement is becoming less and less salient to real conditions in the world.

Let me add that I am no paradigm-shifter. All the paradigm-shifters of my experience seem to be selling snake oil, and that is a very old trade. So far as I can make out, the world is going on as it always has: birth, life, death, diversity, and unity (*e pluribus unum*, indeed!).

There are momentous changes afoot in the world, to be sure, but such changes are always just being born, or in full bloom, or they are dying. Other technological revolutions, other conceptions of knowledge, and other bases of political organization than those we now appear to be giving up, simply constitute the flux of history. Though Heraclitus's river flows on, rural places are major tributaries to the flow, and at this bend of the river they have a lot to tell each other--and much with which to inform all the global flows.

Endnotes

1. Agriculture, however, *must* still mean "rural," since large urban hydroponic wheat and corn farms to supply the global grain market have yet to emerge, though it would be an interesting conceit on which to base a science fiction novel.

2. Recall that these points apply to a century-and-a-half of school improvement. There are plenty of recent exceptions, perhaps an increasing number, and the fact that we are having this conversation is perhaps evidence that things are changing.

3. The grimness of outlook is not evident to everyone in education. But ample evidence suggests a declining standard of living, a shrinking middle class, and a proliferation of deskilled jobs in the US. The school establishment exhorts the populace to prepare for well-paid high-tech jobs, but the supply of such jobs is likely to be inadequate to the demand for them.

4. I am, of course, claiming that we need research about a different kind of improvement. There will be objections to this point. Many educators and professional improvers claim that we have done enough research and that extant research already tells us what we need, and it's not more research. We need, in this view, systemic improvement; authentic learning and assessment; we need to honor multiple intelligences and creativity; we should aim at cultivating lifelong learners (seemingly defined as those willing to undertake an interminable round of formal schooling)--all in the name of preparing the nation for the 21st century of global competition. This is a jejune (underfed) adaptation of the logic of nation-building.

5. The setting was a university, and some of the participants to the conversation seemed prone to dismiss ruralness as a "demographic," and therefore superficial, feature of reality. In general, in both higher and lower education, *rural* connotes an equity issue of not quite pressing concern. The mental image of bucolic poverty, however, is mostly an oxymoron.

6. Educators, especially educational administrators, are apt to associate the distinction between local and cosmopolitan with Alvin Gouldner, who used the terms to describe organizational inhabitants--cosmopolitans were careerists not particularly committed to the organization, but to a larger professional purpose. My usage here sweeps up history, culture, and politics.

7. Jacobs (1984) is a refreshing exception among the social sciences. Her critique of macroeconomics is based on the thesis that cities, and not nations, are the engines of economic growth and health. Macroeconomics, in Jacobs's view, is a fiction based on the need to protect and reinforce national sovereignty. Macroeconomics is a poor tool of economic understanding and management because it substitutes national sovereignty from the local integrity that powers economic life.

8.Though this is a sharp charge, this sort of 'ignorance' (really a form of knowledge) is the basis of the widespread longing for so-called interdisciplinary studies, and the longing for wholism more generally, a longing that actually is not new, but which coincides with modernism. You need to turn to good fiction, however, to follow these longings through the course of the 20th century. This reading is never part of a researcher's training.

References

Anderson, C. (1993). Prescribing the life of the mind: An essay on the purpose of the university. Madison, WI: The University of Wisconsin Press.

Barzun, J. (1959). The house of intellect. New York: Harper & Row.

Benavot, A., Cha, Y., Kamens, D., Meyer, J., & Wong, S. (1991). Knowledge for the masses: World models and national curricula, 1920-1986. American Sociological Review, 56, 85-100.

Consortium for Policy Research in Education. (1996). Public policy and school reform: A research summary. Philadelphia: Author.

Gouldner, A. (1958). Cosmopolitans and locals: Toward an analysis of latent social roles, I. Administrative Science Quarterly, 3, 281-306. (work on the "new class")

Gouldner, A. (1957). Cosmopolitans and locals: Toward an analysis of latent social roles, II. Administrative Science Quarterly, 2, 444-480.

Hofstadter, R. (1959). Anti-intellectualism in American life. New York: Columbia University Press.

Howley, C., Howley, A., & Pendarvis, E. (1995). Out of our minds: Anti-intellectualism and talent development in American schooling. New York: Teachers College Press.

Jacobs, J. (1984). Cities and the wealth of nations: Principles of economic life. New York: Random House. (Excerpt)

Kemmis, D. (1990). Community and the politics of place. Norman, OK: University of Oklahoma Press.

Lasch, C. (1995). The revolt of the elites and the betrayal of democracy. New York: Norton. (review)

McMichael, P. (1996). Globalization: Myths and realities. Rural Sociology, 61(1), 25-55.

Meyer, J., Tyack, D., Nagel, J., & Gordon, A. (1979). Public education as nation-building in America: Enrollments and bureaucratization in the American states, 1870-1930. American Journal of Sociology, 85(3), 591-613.

Mills, C. (1959). Grand theory [fun with Talcott Parsons]. In Mills, C., The sociological imagination (pp. 25-49). NY: Oxford University Press. (excerpt from *The Power Elite*).

Orr, D. (1995). Reruralizing education (draft manuscript). In W. Jackson and W. Vitek (Eds.), Home territory. New Haven, CT: Yale University Press.

Sashkin, M., & Egermeier, J. (1993). School change models and processes: A review and synthesis of research and practice. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, Programs for the Improvement of Practice.

Theobald, P. (1995). Call school: Rural Midwest education to 1918. Carbondale, IL: University of Southern Illinois Press.

Appendix

These books are not the best nor the only works to commend. Many very useful, shorter works specifically devoted to rural education should be consulted. One can, however, easily catch up particular issues in that more specific literature. The *other sort* of reading, merely hinted at below, is much more difficult to prescribe, but much more essential, and, I would argue, a necessary condition of more specific reading.

Literature, Philosophy, Sociology

Arendt, H. (1959). *The human condition.*

Bell, D. (1976). *The cultural contradictions of capitalism*.

Bellah, R. and colleagues. (1985). *Habits of the heart*. (critique of)

Berry, W. (1974). *The unsettling of America*.

Forster, E. M. (1921). *Howards end* [online text]. (film)

Foucault, M. (1979). *Discipline and punish: The birth of the prison*.

Gaventa, J. (1980). *Power and powerlessness*. [no web ref found 4/19/97]

Kemmis, D. (1990). *Community and the politics of place*.

Lasch, C. (1991). *The true and only heaven: Progress and its critics*. [no web ref found 4/19/97]

Lasch, C. (1979). *The Culture of Narcissism*.

Nearing, H., & Nearing, S. (1970). *Living the good life*.

Schumacher, E. (1973). *Small is beautiful*. (E.F. Schumacher Society)

Stegner, W. (1946). *Second growth*. (novel)

Veblen, T. (1899). *The theory of the leisure class*. (online text)

Weber, M. (1904). *The Protestant ethic and the spirit of capitalism*. (online excerpt)

Walker, M. (1946). *Winter wheat*. (novel)

West, C. (1989). *The American evasion of philosophy: A genealogy of pragmatism*.

Whisnant, D. (1983). *Modernizing the mountaineer*.

Education

Brown, R. (1991). *Schools of thought: How politics shape literacy in the classroom*. (review)

Carnoy, M., & Levin, H. (1985). *Schooling and work in the democratic state*.

Counts, G. (1930). *The American road to culture*. (related analysis)

Cubberley, E. (1922). *Rural life and education*.

DeYoung, A. (1989). *Economics and American education*.

DeYoung, A. (1995). *The life and death of a rural American high school*. (review)

Gatto, J. (1990). *Dumbing us down*. (for the skeptical: placed in context)

Jencks, C., and colleagues. (1979). *Who gets ahead*.

Katz, M. (1968). *The irony of early school reform.* (recommended by Noel McGinn)

Keizer, G. (1988). *No place but here.* [no web ref found 4/19/97]

Kliebard, H. (1986). *The struggle for the American curriculum.* [no web ref found 4/19/97]

Nachtigal, P. (1982). *Rural education: In search of a better way.*

Shea, C., Kahane, E., & Sola, P. (1989). *The new servants of power.*

Sher, J. (1977). *Rural education: A reassessment of the conventional wisdom.* [no web ref found 4/19/97]

Spring, J. (1976). *The sorting machine: National educational policy since 1945.* (other work)

Tyack, D. (1974). *The one best system: A history of American urban education.*

Wigginton, E. (1985). *Sometimes a shining moment.* [no web ref found 4/19/97]

About the Author

Craig Howley, Director
ERIC Clearinghouse on Rural Education and Small Schools
Appalachia Educational Laboratory
Phone: 800-624-9120 email: howleyc@ael.org

http://www.ael.org/erichp.htm

I've written about, studied, and lived in rural places. (It's debatable whether or not I still live in a rural place, but the local chamber of commerce says I do, given that our house sits 2 miles north of I-64).

Culture, politics, economics, and history concern me. I wish schools were better at promoting 'the life of the mind' (whatever that is; finding out is part of the adventure) among everyone. And I think there are reasons they don't, but these reasons constitute more than just inattention or foolishness. Culture, politics, economics, and history suggest reasons.

Literature (fiction) may be a much better guide to true education in rural places than the sorts of poor studies we educationists sponsor. Check out Wendell Stegner's *Second Growth* (circa 1950) or Annie Proulx's *The Shipping News* (circa 1990) and even E.M. Forster's *Howards End* (circa 1920). These folks have preserved something we have tried desperately to abandon, but can't actually escape. The wonder is that, though these books (and many more) treat the dilemmas of rural life, they also deal with the idea of a true education more universally. Now, that's fun because it's not easy. In particular, novels don't lend themselves to translations as cookbooks.

Teaching well is the most difficult work in the world. We make a great mistake with attempts to make it easy or happy. Happiness is not a worthy aim for education, nor is getting and holding a good job.

Copyright 1997 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://olam.ed.asu.edu/epaa

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey@olam.ed.asu.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin
John Covaleskie
Northern Michigan University
Alan Davis
University of Colorado, Denver
Mark E. Fetler
California Commission on Teacher Credentialing
Thomas F. Green
Syracuse University
Arlen Gullickson
Western Michigan University
Aimee Howley
Marshall University
William Hunter
University of Calgary
Daniel Kallós
Umeå University
Thomas Mauhs-Pugh
Rocky Mountain College
William McInerney
Purdue University
Les McLean
University of Toronto
Anne L. Pemberton
apembert@pen.k12.va.us
Richard C. Richardson
Arizona State University
Dennis Sayers
University of California at Davis
Michael Scriven
scriven@aol.com
Robert Stonehill
U.S. Department of Education

Greg Camilli
Rutgers University
Andrew Coulson
a_coulson@msn.com
Sherman Dorn
University of South Florida
Richard Garlikov
hmwkhelp@scott.net
Alison I. Griffith
York University
Ernest R. House
University of Colorado
Craig B. Howley
Appalachia Educational Laboratory
Richard M. Jaeger
University of North Carolina--Greensboro
Benjamin Levin
University of Manitoba
Dewayne Matthews
Western Interstate Commission for Higher Education
Mary P. McKeown
Arizona Board of Regents
Susan Bobbitt Nolen
University of Washington
Hugh G. Petrie
SUNY Buffalo
Anthony G. Rud Jr.
Purdue University
Jay D. Scribner
University of Texas at Austin
Robert E. Stake
University of Illinois--UC
Robert T. Stout
Arizona State University

---

archives | abstracts | editors | board | submit | comment | subscribe | search

Education Policy Analysis Archives

Volume 5 Number 13 **June 10, 1997** **ISSN 1068-2341**

A peer-reviewed scholarly electronic journal.
Editor: Gene V Glass Glass@ASU.EDU.
College of Education
Arizona State University,Tempe AZ 85287-2411

Copyright 1997, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

Qualitative Research Methods:

An essay review

Les McLean
Margaret Myers
Carol Smillie
Dale Vaillancourt
Ontario Institute for Studies in Education/University of Toronto

Miller, Steven I. & Fredericks, Marcel (1994) *Qualitative Research Methods: Social Epistemology and Practical Inquiry.*
New York: Peter Lang. 159 pages.

> …the more we try to unpack the notion of evidence for the sake of clarity, the more problematic it becomes, especially in trying to provide adequate justifications for educational policy making issues. (p. 119)

Abstract

The authors ask us to explore the topic of "qualitative confirmation" in relation to the processes and outcomes of qualitative research practice. The question that directs their inquiry is "how can we make a case that qualitative data or findings warrant the inferences about the topics we are studying?" We review the historical discussion of confirmation theory within the logic of discovery, consider hypothesis generation and methodological decisions as instruments of the research process and then apply the Miller and Fredericks framework of rules to a published report of qualitative research (Glass, 1997). Full bibliographic references may be viewed by clicking on References (below) or on one of the linked citations in the text. We end our review with an appreciation of the work.

Purpose
Description
Related Works
Rules for confirmation
Application of rules
Appreciation
References
Notes

About the Authors

Les McLean

Les McLean is a Professor (Emeritus, as of July 1, 1996), Ontario Institute for Studies in Education of the University of Toronto (OISE/UT). He received his doctorate in Educational Psychology from the University of Wisconsin in 1964, specializing in statistics and research methods. After teaching at Columbia University Teacher's College, Dr. McLean joined the OISE faculty in 1966, teaching courses in measurement, statistics, quantitative and qualitative research methods and program evaluation. He continues to teach on a contract basis, year-to-year. Research and development projects have included a national study of the evaluation of student achievement for the Canadian Education Association, several province-wide surveys of student achievement, direction of Ontario's participation in the Second International Mathematics Study and research into mathematics/language relationships in curriculum and pedagogy. Les, Doris Ryan and Barbara Burnaby directed an evaluation of four projects in the Canada/China Human Resources Development programme for the Canadian International Development Agency. Publications include, *The Craft of Student Evaluation in Canada* (Toronto:CEA, 1985), *Learning About Teaching from Comparative Studies* (Toronto:Min. of Educ., 1987, with Richard Wolfe and Merlin Wahlstrom), and "The U.S. national assessments in reading: reading too much into the findings" (*Phi Delta Kappan*, 69, 5, 1988, 369-372, with Harvey Goldstein). "Time to replace the classroom test with authentic measurement" (*Alberta Journal of Educational Research*, 36(1), 1990, 78-84), "Student evaluation in the ungraded primary school: The SCRP principle" (*Proceedings of the Second Canadian Conference on Classroom Testing*, D. Bateson, Ed., UBC, 1992) and "Pedagogical relevance in large-scale assessment" (*Advances in Program Evaluation*, R. Stake, Ed., 1991).

For recent publications,see http://www.oise.utoronto.ca/~lmclean/>.

Les McLean, Professor-Emeritus lmclean@oise.utoronto.ca
Measurement and Evaluation (416) 923-6641, ext 2478
Department of Curriculum, Teaching and Learning
OISE/Univ. of Toronto
252 Bloor Street West
Toronto, Ontario, CANADA M5S 1V6

Carol Smillie

Associate Professor Dalhousie University School of Nursing. Vice President Canadian Cancer Society. Board of Directors of the National Cancer Institute of Canada and the Canadian Cancer Society. Director of the Sociobehavioural Research Satellite Centre of the National Cancer Institute's Sociobehaioural Research Network. Practice of Nursing has been concentrated in the area of Community Health and working with the Volunteer Health Sector.

Margaret Myers

Trained as Registered Nurse at St. Martha's School of Nursing, Antigonish, Nova Scotia; Bachelor of Science in Nursing from University of Western Ontario; Masters of Arts in Education from St. Francis Xavier, Antigonish, Nova Scotia; Have practiced in a variety of settings, including Nursing Education and Nursing Management. Speciality area is Program Developement and Implementation. Chairperson of the Diabetes Nursing Interest Group with the Registered nurses Association of Ontario. Am currently writing a textbook on diabetes education for health care professionals. Enrolled in Doctoral program in Curriculum at OISE at the University of Toronto. Married and living in London (Ont.)

Dale Vaillancourt
Dale's main interest is in using research to enhance online teaching and learning experiences, especially for women entrepreneurs. She is using her 25 years in the telecommunications industry and a full-time academic studies program in computer applications to help her meet this goal. For further details, see: http://www.oise.utoronto.ca/~dvaillancourt/resume.htm

---

Copyright 1997 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://olam.ed.asu.edu/epaa

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey@olam.ed.asu.edu .**EPAA Editorial Board**

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Marshall University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Rocky Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina--Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher Education

Mary P. McKeown
Arizona Board of Regents

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Richard C. Richardson
Arizona State University

Anthony G. Rud Jr.
Purdue University

Dennis Sayers
University of California at Davis

Jay D. Scribner
University of Texas at Austin

Michael Scriven
scriven@aol.com

Robert E. Stake
University of Illinois--UC

Robert Stonehill
U.S. Department of Education

Robert T. Stout
Arizona State University

---

archives | abstracts | editors | board | submit | comment | subscribe | search

Purpose of our Review

The purpose of our review is to engage the reader in further discourse on this topic of qualitative confirmation. We will use Miller and Fredericks' (M&F) work as a vehicle to initiate debate and demonstrate its usefulness for critiquing published qualitative research studies.

Review Method

Following a review of the historical roots of confirmation theory and a description of M&F's method for establishing qualitative confirmation, we will apply the method to a journal article by Sandra Rubin Glass (1997), an extensive investigation of teacher and principal autonomy in both private and public schools. Her article is an attractive example because she offers readers ready access to much of her original data.

Historical Roots of Qualitative Confirmation

From a historical perspective, M&F take advantage of the *logic* provided by positivist theories of confirmation (e.g., Carnap, 1962, Hempel, 1965; Swinburne, 1973) and apply it to qualitative research. The authors argue that the use of a hypothesis and numerical probability statements continues to lend weight and credibility to conventional research and they seek to persuade us to transfer elements of the positivist paradigm to qualitative research. They recognize that "confirmation theory" involves complex issues and its application in the postpositivist arena is not a simple matter. Swinburne(1973) claimed that confirming one's findings must be reduced to probability (computational) statements, but M&F suggest that reducing qualitative research to quantitative calculations would overlook the advantages that naturalistic inquiry has to offer. From Carnap (1962), the term "confirmation" involves issues of classification (what is evidence), quantity (how much is good evidence) and comparison (what is the degree of firmness among the evidence).

In M&F's eyes, this raises important questions: whether numerically-based evidence is an option and whether the term confirmation can have different connotations. From Hempel's (1965) constitutive and regulative rules of confirmation, M&F recognize that hypotheses can be further supported by propositions if the propositions are defined legally and/or operationally. Again borrowing from the positivist school, issues related to indeterminacy and incommensurability (Quine, 1963, p.5,9) are introduced. If there is no objective reality and several data collection methods are allowed, triangulation is just as likely to yield contradictory views as it is to produce supportive ones. Which translation manual (2) should one use (indeterminacy)?

Similarly, if competing translation theories inform the observations, and therefore cannot logically be used to determine which theory is correct (incommensurability), what does one use to make a decision? Charting the logic underlying one's inquiry process lends validity or correctness to the inferences (see also LeCompte & Preissle, 1993). They do not require translation manuals congruent with one's epistemological and ontological assumptions, but instead prompt the researcher to outline what other translations are possible and determine which one or ones would be valid and why. Most qualitative researchers select their preferred method of inquiry, make a general statement about the data and show a few excerpts; this does not mean, however, that the selected process or the inferences are valid. M&F believe that in every study there is an *a priori* or *a posteriori* hypothesis, whether explicitly stated or nor, and that there can and should be a demonstration of how one's epistemological and ontological assumptions, triangulation approach, translation manuals and weighing of evidence link the evidence to that hypothesis. They adapt Hempel's rules to qualitative research (Hempel , 1965) and introduce the term, "qualitative confirmation."

Definition of Qualitative Confirmation

M&F define the term as: "those logical conditions that must obtain between the evidence and hypothesis" (p. 11). The authors admit that the term is an "oxymoronic-sounding label" (p.1). However, they argue that their adaptation of the Hempel rules helps decide whether the logic underlying the methods chosen and the sampling and the weighing system for the evidence satisfies the conditions of qualitative confirmation. The rules are also expected to help one minimize the influence of personal bias. M&F recommend a systematic way of conducting and critiquing qualitative research that makes explicit the researcher's process from start to finish and highlights where pitfalls can be prevented. As already noted, one begins with an a priori or a posteriori hypothesis, which may be a transformation of the research question. This emphasis on a hypothesis would seem to set M&F apart from other qualitative research theorists (e.g., Glaser and Strauss, 1967; Stake, 1995), but we think not. See, for example, Strauss and Corbin (1990) on "The Research Question" (pp. 36-39), and Stake, Chapter 2 (pp. 15-33).

The overt thinking that guides the process and the resulting evidence-instances (3) will ultimately be used to support or reject the hypothesis. This act of specifying one's logic, guided by rules, touches all phases of the qualitative research process; it is a demonstration why the planned activities and resulting evidence support the hypothesis. M&F state: " In other words, qualitative data have to 'demonstrate' their utility as potential evidential candidates for confirmation. It is not simply a matter of amassing positive evidence-instances for a hypothesis, but also showing *why* they contribute to supporting a given hypothesis" (p. 33, emphasis in original). Thus, qualitative confirmation provides a way of reporting the researcher's thought processes, helps promote a constructivist approach to qualitative research and answers the demand for increased rigor that Miles and Huberman (1984) identify: "Despite a growing interest in qualitative studies, we lack a body of clearly-defined methods for drawing valid meaning from qualitative data. We need methods that are practical, communicable, and not self-deluding; scientific in the positivist's sense of the word, and aimed toward interpretive understanding in the best sense of that term" (p. 21).

Return to the Contents Table.

Go on to the next section.

Description of the Book

This is a scholarly book, in conception and content (but alas not in execution). We mean by "scholarly" that it digs deep for core concepts in what has been written about qualitative research methods-ideas, constructs, images that help us understand more than one case study, more than one ethnography, more than one narrative. Do such "core concepts" exist? Miller & Fredericks (M&F) think so. Otherwise how could they organize a book around the construct "qualitative confirmation"? What could it possibly mean to "confirm" findings from qualitative research? Reading Smith and Heshusius (1986) or Wolcott (1994), one might doubt the possibility.

Then there is postmodernism: doubt, doubt, doubt; everywhere we look there is doubt. A few years ago, the head of the British Museum of Natural History reviewed challenges to Darwinian theories of evolution and found merit in some of them. Attacked by critics for sowing doubt, he replied, "Doubt is splendid stuff." (4) Doubt may indeed be splendid stuff, but we humans seem to crave something else, something more solid in our mental life--something such as confirmation.

M&F offer material that is relevant to researchers everywhere but their editors have done them a disservice by presenting the work in a careless and awkward way. The lack of bold faced type and subheadings make an already difficult text even more difficult to read. References are given at the end of each chapter and each appendix, except for Chapter Three which has none. As a result, several are missing and there are many duplications. Outright errors appear that would be caught by even a cursory glance from an editor. (5) A curious feature of the book is that of its 153 pages, 62 are devoted to three appendixes which contain crucial elaboration or explanation. That said, we believe these authors provide a unique and powerful contribution to qualitative research--concepts and procedures that are worthy of study. *Confirmation* is the central concept of the book, a concept M&F argue, and we agree, has not been adequately addressed within the qualitative field.

Chapter One

Confirmation theory is introduced as the "major issue in qualitative inquiry", defining it as a description of what it means to say that the evidence relates to the hypothesis in the qualitative process. It is "the theory of when and how much different evidence renders different hypotheses probable" (Swinburne, 1973, p vi). It is acknowledged that this process is not free from assumptions and thus evidence can ultimately only be known through the senses. What constitutes data depends largely on who is looking for it; where mathematics may deal with pure fact, physics may not. The concept of weighing the evidence is introduced and *weighing* is distinguished from *weight*. Weight refers to the outcome while the weighing is the process of arriving at the outcome. In addition, weight can only be assessed by giving a rationale for the weighing of the evidence. Although M&F concede the concerns that have arisen over Hempel's classical framework of rules (1965), they value it and build directly upon it. Hempel defined: hypothesis, observation report, observation sentence, entailment, direct confirmation, confirmation and development of a hypothesis for a class of individuals. (quoted by M&F on pages 8 and 9). They also cite Carnap (1962), who argued that although confirmation is central to understanding the evidence relationship, the term is also susceptible to different interpretations. Carnap, however, relied on computational interpretations and M&F argue that that would defeat the purpose of qualitative confirmation. The authors' intent is to use a non-probability framework to focus on a logic of discovery while not excluding the possibility of validation.

Since evidence comes in so many forms, rules are needed to help assess the evidence in relation to the research question (hypothesis). It is also necessary to develop a series of related but separate rules which bear only on the characterization of the data itself (p.12). M&F suggest that the ways qualitative researchers make a case for their findings will depend on two major aspects: the *labeling of* and the *justification for* using the rules. Two categories of rules are required: *constitutive* (defining what counts as a social situation or practice) and *regulative* (prohibiting or prescribing actions in situations defined by constitutive rules-taken from Greenwood, 1989).). In their discussion, M&F are "trying to understand what it means to be 'rational' in pursuing the activities of qualitative inquiry" (p. 14), and they reach back to "camps" defined by Wittgenstein, Popper and Donald Davidson (1984) for insight into rationality. To the present reviewers surprise, they then add "sociologists" to the list, a camp that "tries to determine the extent 'natives' hold their beliefs even in the face of (supposedly) other more 'rational' beliefs" (p. 14). Sociologists?

The last "camp" specializes the thorny and recurring issue of "translation", an issue that permeates any discussion of qualitative research. Whenever we use words to communicate the behaviour and/or feelings of others, the communication involves translation even when we all seem to be speaking the same language. Translation is discussed in every chapter of M&F, from many perspectives. Translation issues revolve around the ideas of indeterminacy and incommensurability. These two constructs are not dealt with at length. The "indeterminacy thesis" is: "because theories of meaning (i.e. those referring to natural languages) are not concerned with 'the fact of the matter', as are scientific theories in the natural sciences, it is possible to derive multiple translation manuals which may be incompatible but adequate for interpreting the behavior in question." The "incommensurability thesis" suggests that there is not even one translation manual that is adequate, as when trying to choose between two competing theories, for example (p.15). (For a comprehensive discussion of these concepts see Quine, 1960.) Translation is often at the level of data analysis rather than confirmation. Rationale for data collection processes and the subsequent description or translation of the data often is directed by the chosen research methodology. M&F close Chapter One with six statements describing how the above considerations relate to the notion of qualitative confirmation.

Chapter Two

The second chapter is devoted to the subject of hypotheses in qualitative research, not surprisingly since hypotheses are central to their entire approach. Hypotheses are viewed broadly "as statements which direct inquiry, in a Deweyan sense, in relation to a theoretical framework, but without the necessity of such hypotheses being strictly deducible from such a framework" (p. 21). A priori hypotheses may emerge where the researcher has a theoretical framework in mind, for example, beginning with a category of behavior and linking it to a category of individuals. In contrast, the researcher may begin with an interest in a category of individuals and subsequently seek to link them to a category of behavior. M&F argue that both are qualitative because they are formulated in the context of a qualitative study and because their confirmation is dependent on qualitative data; they dodge the obvious tautology by acknowledging that both types can be formulated as well in quantitative studies. Confirmation, they say, is central to all research activities, but in qualitative research it is based on plausible logical relations which apply first to the data and then to the hypothesis. Qualitative research seeks to understand human behavior "from a perspective which, methodologically, requires qualitative data" (p. 25). Rules, they say, are needed if a case for qualitative confirmation is to be made. Since rules must apply both to data and hypothesis, M&F undertake to define what they mean by data-and their definition is entirely conventional: field notes, interview data, historical accounts and the like (pp. 26-7). The term "evidence-instance" is introduced to describe a discrete item of information to be put forward as part of a confirmation argument. A long elaboration on the meaning of "evidence" is presented in Appendix B, including more discussion how data become evidence. (6)

They discuss triangulation, breaking no new ground, and then tackle the really tough question: what counts as evidence and how do you judge its importance? Nothing new here, and we are all reassured to read, "It becomes quite difficult, then, to develop hard-and-fast rules for determining appropriate 'weighting' procedures for these kinds of issues" (p. 30). Some considerations that "may be helpful" include:

- The weight of the data-evidence is not necessarily synonymous with the "amount" of the data-evidence. The term "weight" can be loosely translated as an "absence of negative cases," that is, disconfirming instances. ...
- Where the term "amount" of evidence is used, the implication is that some of the evidence must consist of (negative) disconfirming instances. ...

M&F assert that contrary findings do not necessarily disconfirm the original hypothesis, for they may serve to clarify a previously unknown dimension. The matter of confirming vs. disconfirming instances is discussed more fully in Chapter Three. They tell us that using more than one data set can enhance the validity or reliability of data if it can be shown that the data sets are relevant to the problem--however, they neglect to explain how this can occur. M&F relate the purposes of using a triangulation process but the inherent problems associated with it are listed and not explained. We recommend that those unfamiliar with the process of triangulation consult other sources.

In what must have been an afterthought, the final paragraph of the final section, "Conclusions", introduces a muddled discussion of the terms "relationship" and "association". Hopes that the muddle will be cleared up are not fulfilled.

Chapter Three.

In chapter three M&F present their four rules for qualitative confirmation and instruct us in the use of these rules, including how to handle negative cases or cases of disconfirmation. Overall, it is the most carefully constructed and clearly presented chapter. How qualitative findings may become evidence for a hypothesis or research question is the focus of the discussion here and the authors impress us with how their qualitative research rules apply to both sides of the research continuum. One advantage claimed for researchers is that the rules allow for the development of qualitative research studies with confirmation in mind. They also suggest how research findings are to be interpreted when choosing data sets. As for qualitative disconfirmation, some additional advice is given in the interpretation of negative evidence-instances. Researchers usually have to use the qualifier "some" instead of "all" (Quine-Duhem Thesis) because scientific theories may be composed of auxiliary hypotheses for certain predictions (p.48). Auxiliary hypotheses can allow incompatible evidence while still not disconfirming the original hypothesis. The null hypothesis also finds a place in Miller and Frederick's qualitative confirmation theory.

Chapter Four.

Chapter Four moves into the practical application of rules to five articles published since 1980. M&F pick up on a suggestion from Miles and Huberman (1984) that researchers ask systematically a series of questions before, during and after the research: (p.53): . what is (are) the major research question(s) or hypotheses? . what method(s) will be used and what sort of data yielded? will other(indirect) data sources be utilized? . how will the case be made that the research question/hypothesis has been or has not been confirmed? . if mixed data sources are used, how will they be handled in terms of confirmation/ disconfirmation? All researchers are advised to include a final section in the research report (possibly as an appendix) which explicitly addresses the issue of qualitative confirmation. Such a section helps the researcher focus on the larger epistemological-theoretical questions raised by the study and reminds us that the presentation of our findings can be regarded as a type of translation manual as well as a demonstration of the rigor in one's research.

The application of confirmation theory to the five caseshelps a little to understand M&F's method, but their choice of articles left much to be explained. A graduate student listed 20 reports of qualitative research from the years 1970-1990. The same student selected 7-10 for further study, in an attempt to present a "rough cross section" of work. In our view, a more qualitative approach would have been preferable: purposive choice to illustrate application of the rules rather than this version of sampling. In only one study, for example, was there any attempt to discuss negative or disconfirming instances. In most there was no attention to the issue of confirmation, but in one or two there was an implicit suggestion that confirmation had been achieved. There was no discussion of the weighing the evidence in making qualitative claims. Given the rather radical nature of M&F's proposals, it should perhaps come as no surprise that research conducted and published more than a decade ago does not adhere closely to the Rules. In order to help future researchers, M&F provide a "Checklist for Qualitative Confirmation" at the end of the chapter, including the questions mentioned above and ending with "What, after all, has one discovered/concluded from this investigation? Is it truly warranted?" (p. 66-7)

Chapter Five.

The final chapter (before those three long appendixes, that is) is entitled, "Epistemological Asides and Conclusions"; fearlessly confronting the big question of "truth". Suppose a study has met all the criteria for qualitative confirmation, "Is this all that is necessary for declaring that the findings, then, are 'true'?" (p. 73) M&F advise us to look at various philosophical camps, bringing

Winch again (1958, 1964) and again invoking the "camps" (from Chapter One). Qualitative confirmation provides a framework for establishing qualitative data as evidence for the hypothesis, that is, it meets the requirement of rationality, but it is less successful as regards indeterminacy. It is not possible to rule out different or competing translation manuals.

As readers will have gathered by now, this concept of "indeterminacy" pervades the book--as it does the field of qualitative research. It is a way the philosophers have discussed whether there can be any "truth" and the way M&F discuss the prevailing view in qualitative circles that there is no truth-there are no methods that allow us to reduce indeterminacy to zero. Readers may feel, with justification, that M&F's discussion of the concept is fragmented and incomplete. Fear not; Appendix A (24 dense pages) is devoted to it.

Appendix A. Some notes on the nature of methodological indeterminacy.

Here the authors return to the intellectual roots of qualitative confirmation and elaborate the construct, "methodological indeterminacy". M&F revisit Quine and lean heavily on Roth's 1987 book, Meaning and method in the social sciences: A case for methodological pluralism. Roth, echoing Winch (1958), opens his book by referring to "the general collapse of positivism" and concludes (following Quine) that multiple translation manuals are not only possible but inevitable. There is no "fact of the matter" in relation to the human sciences.

Where can we go from here? Undaunted, M&F press on, seeking to dodge the philosophical bullet by concentrating on practical concerns, namely methods.

> ...the genuine problem of indeterminacy for the human sciences does not lie at the level of debates concerning "hermeneutic" vs "scientific empirical" *views* of human action, but rather at the level of *specific* methodological techniques that ultimately are the constitutive elements for determining the existence (or lack of it) of indeterminant translations. (p. 93, emphases in original)
>
> ......
>
> To be clear, we are not arguing that the definition of methodology as the application of specific techniques and procedures is sufficient to resolve indeterminacy in all situations, but only that it is a necessary condition both to establish its existence and to demonstrate *possible* ways of reducing it. (p. 94, emphasis in original)

Some clarity does emerge: specific techniques and procedures (methods) are needed to *establish the existence of* indeterminacy. It is always there, we suppose, but until you apply the techniques and procedures (collect the data?), you don't see it Now that positivism has collapsed, we go farther: application of methods "may *generate* a type of indeterminacy" (p 101), or "types of indeterminacy *that are produced by* competing methodologies" (p. 103, emphasis added). This is surely correct, because different methods bring to light different amounts and kinds of indeterminacy-hence the term, "methodological indeterminacy". M&F borrow from quantitative methods and discuss differences "between" and "within" methodological approaches, with regard to both qualitative and quantitative approaches. The discussion of qualitative approaches is much more insightful than that of the quantitative, the latter naïve and verging on the trivial. We question M&F's assertion, citing Fuller (1988), that "there is a heavier 'burden of proof' on the qualitative side of the equation for showing either indeterminacy or lack of it." (p. 105) A better statement, in our opinion, is that for practicing researchers, "indeterminacy is usually perceived itself as a 'working hypothesis'; one which may never in principle be unequivocally 'accepted' or 'rejected' but one, nevertheless, that is capable of empirical inquiry." (p. 105)

Appendix A concludes with a short discussion of the status of theory in the human sciences. (M&F leave no topic completely untouched!) They cite a belief among "some quantitative researchers" that improvements in measurement will solve our problems, including give us better theories. In our opinion, they have it backwards: better theories lead to better measures (and the process recycles). Let M&F have the last word, with which we completely agree:

On the other hand, for those who do not see the possibility (or usefulness) of equating the human sciences with the natural sciences, the indeterminacy reflected by methodological applications can be "reduced" over time, but its reduction is directed towards making human behavior more "intelligible" rather than more "scientific". (p. 108)

Appendix B. Clarifying the "adequate evidence condition" in educational theory and research

The authors set out to convince us how important it is to clarify what they call the "evidence condition". Lack of attention to what is meant by evidence "has resulted in a restricted view of this term with a correspondingly unwarranted optimism regarding the formulation, implementation and evaluation of educational policies and practices." (p. 117) Faithful to their book title, they promise a thick description type of analysis to attain their goal. If "dense" is the same as "thick", then they deliver on this promise.

Israel Scheffler (1965) is cited early, as we would expect, and not just because he introduced the phrase, "adequate evidence condition". No discussion of evidence in educational research would be complete without at least a mention of this classic. New sources to us are Lakoff (1987) *Women, fire, and dangerous things: What categories reveal about the mind*, and Lakoff & Johnson (1980) *Metaphors we live by*, insightful discussions of the nature and process of category creation. Since qualitative researchers live and die by their categories, these are helpful references for anyone (such as the authors of this review) who had overlooked them. The process of categorization is the formation of "Idealized Cognitive Models …culturally unique and semantically-based modes of perception and reasoning whereby users of a language construct cognitively-based models to explain social and physical reality." (p. 124) There are five types of models: Cluster, Metonymic, Social Stereotype, Ideal and Metaphoric.

We are not attempting a summary of this appendix. M&F describe the concept of evidence as having "entrenched ambiguity" and "overall complexity", and after reading the appendix several times we heartily agree. Lawyers and judges struggle with it daily, of course, and law schools have whole courses on it. M&F also write, "Paradoxically, however, the more we try to unpack the notion of evidence for the sake of clarity, the more problematic it becomes, especially in trying to provide adequate justifications for educational policy making issues." (p. 119) Here are a few points to give the flavour of the work.

Among the ambiguities cited is the "conflation" of the terms evidence and data, as we noted in the description of Chapter Two and in Note (6). According to M&F, data are *candidates* for evidence but only become evidence "to the extent that they are consistent with both the larger class of rules identified with the particular methodological approach, and with the rules regulating the application of a particular technique or tool within the methodological approach." (p. 118) Whew!

Educational implications are offered, including ambiguities surrounding the concept of "school effectiveness" and the lack of consensus on adequate evidence for effectiveness. You might not agree with everything they write, but you will be led to think hard about it.

Appendix C. Reciprocal paradigm shifts and educational research: A further view of the quantitative-qualitative dilemma

In the concluding appendix, M&F state their belief and concern that while those involved in educational research perceive a paradigm shift from quantitative to qualitative inquiry, many are not convinced that the qualitative-interpretationist view can be sustained on the basis of conventional scientific criteria. Some, Miles and Huberman (1985) for example, manage to sidestep this issue and some recognize qualitative research but believe that the eventual testing or proof of causality or theoretical prediction must be left to scientific empiricism (Popper, 1969, Rosenberg,1988). M&F concentrate their discussion on the interpretationist framework and they examine some commonly held assumptions that underlie the two research processes. They present their position that quantitative and qualitative approaches need not be viewed as contradictory in terms of some educational problems within some contexts, and a methodological mix is not only possible but desirable.

Miller and Fredericks do not agree with the asumption that scientific empirical data is quantitative data and that the constructs of validity and reliability are crucial issues only in quantitative studies.

One of our central claims will be that not only are the crucial issues of reliability and validity perfectly general across all research based forms of inquiry; but also that in meeting the requirements of reliability and validity, interpretationist accounts may yield warranted conclusions that, while different in form and content from scientific empirical claims, can, nevertheless, be compatible with them. (p. 137)

They remind us that these assumptions have their roots in the traditional confirmation theory of Carnap (1962) and Hempel (1958) and are drawn from a "probabalistic-operational" perspective. This approach to the definition of scientific evidence seems to rule out the possibility of interpretationist accounts of data that can only be understood rationally as qualitative data. In response, they put forward and discuss the following assumptions (p. 139):

1. interpretationist accounts within the human sciences are correctly based on the direct or presumptive use of non-quantitative research strategies,
2. these strategies must be viewed as being fundamentally different in their qpplication and in the interpretation of the data they produce from those research strategies based on probability assumptions,
3. they are, nonetheless, subject to the general scientific constraints of reliability and intersubjectivity, and
4. they can be employed as alternative means for "confirming" hypotheses and/or providing for critical tests for theoretical perspectives.

These assumptions lead them to confident statements about the tough issues of causality and theoretical interpretation of reports from qualitative studies. "First, the qualitative research agenda stipulates only that causal connections (as either necessary or sufficient conditions) of behavior be meaningful in terms of the native's own accounts and that such connections be accurately described. Secondly, the investigator's theoretical interpretation of these accounts *can* incorporate 'higher order' or more general constructs, but without the implication that their use must *necessarily* 'reduce' the original accounts." (p. 145, emphasis in original)

Return to the Contents Table.

Go on to the next section.

Other views relevant to qualitative confirmation

This section is presented to place M&F within the larger qualitative research community. Our discussion is in no way a complete review of the voluminous literature; it highlights different interpretations of the hypothesis setting process, validty and reliability concerns and understandings of translation and triangulation methods. In this section, we emphasize with italics the words that seem to us to be versions of *confirmation*. In many ways, LeCompte and Preissle (1993) support M&F's concept of qualitative confirmation. LeCompte and Preissle agree, at least metaphorically, that validity is involved in many aspects of the inquiry process: "How validity is defined and treated varies according to what researchers do, what tasks they are undertaking, and in what phase or stage of the research they are in" (p. 325). Theoretical frameworks, general design, context, participants, researcher experience and procedures of data collection and analysis have a bearing on the issue of validity. As LeCompte and Preissle say, "Consequently, although we urge scholars to discover and formulate what their research philosophy is, we believe that it is only one factor contributing to how validity is defined" (p. 326). They also caution, as do M&F, that replacing qualitative processes with strictly quantitative ones erroneously prompts a single consolidated definition of validity and potentially jeopardizes richness of detail and creativity. For LeCompte and Preissle, qualitative research is idiosyncratic and data analysis entails an emergent process: "Even midway through an analysis,uncertainty and frustration accompany the unfolding direction" (p. 330). They see qualitative research as loosely connecting researchers who come from a broad spectrum of philosophical traditions; there is not just one. They could ask that all the different qualitative researchers state how their philosophies decide validity and then apply those guidelines to the study, but they stop short of this because they see it as an *a priori* assignment approaching "determinism" (p. 326).

A comparison of their descriptions of research phases demonstrates how they propose to obtain validity (or qualitative confirmation) especially in the areas of: formulating goals, developing a research design, selecting data sources, experiencing and directing the research, collecting data, collaboration, comparing phenomenon and data analysis. Part of developing research questions is to ensure that the research goals, the context of the situation and the interests of the stakeholders are aligned. M&F would say that this process entails what must also be considered when developing a hypothesis. Much of the discussion around validity stems from concerns about the sources that are assumed to provide validity. LeCompte and Preissle argue (as do M&F) that qualitative research methodologies cannot discount the range of ontological and epistemological assumptions and theories that are at their disposal. The key to validity is that researchers must be aware of and acknowledge the use of disparate views. The kinds of evidence colleagues accept as legitimate and adequate may be affected by who is being studied. More data may be required because of who is going to scrutinize the results. Therefore, who or what is to be studied must be considered when deciding on a design. This is similar to M&F's concept of weighing (but not 'weighting').

Will there be other extraneous factors such as background evidence that needs to weigh into the confirmation of the hypothesis? Researcher's background and role in the investigation are central to how the validity is addressed:

> History teaches that attention to the individual researcher is relevant to validity in qualitative research. What background and training does the researcher bring to the investigation? How carefully, thoroughly, openly, and honestly are researchers known to do their work? Who was responsible for the researcher's training? What reputation has the scholar earned in previous investigations? What does the researcher report about participation in the research? Introspective and reflective amounts of influence on what is seen and heard contribute to the audience's confidence that the researcher attempted to track these factors." (p.329).

This is similar to M&F' constitutive and regulative rule that revealing the researcher's training and practice as well as the ethical or operational practices will help us assess how much *credibility* can be given to the researcher's work. LeCompte and Preissle agree with M&F that a systematic way of collecting data can be used to give more credence than one that is not. Both pairs also agree that just because something is done correctly, it may strengthen the research but does not necessarily mean that the results are sufficient to meet the criteria. LeCompte and Preissle suggest many ways for researchers to enhance confidence in their results, for example, through collaborative participation with the participant, congruency between theory and observation, inter-method and inter-observer checks, personal reflection and introspection. Therefore, although LeCompte and Preissle would determine their analysis procedures at a different stage than M&F, they do agree with M&F that several options are available. Both agree that one should use multiple methods to reduce the possibility that bias will affect the credibility or validity of the results.

LeCompte and Preissle do not strive to provide conventional external validity because small sample sizes usually make this task impossible. However, they state that comparing phenomenon is useful and can be achieved by defining the "typicality" (Wolcott, 1973) of a phenomenon. Threats to comparing phenomenon are whatever obstructs or reduces a study's *translatability*. Translatability is the degree to which the researcher can adopt theoretical frames, definitions and research techniques accessible to and understood by other researchers in the same or related disciplines. Thus LeCompte and Preissle agree with M&F that alternative translations should be considered. However, LeCompte and Preissle think of translatability in terms of its usefulness in linking with others, whereas M&F think of translatability for reducing bias and confirming evidence or a hypothesis.

The one other area in which LeCompte and Preissle are distinct from M&F is in data analysis. LeCompte and Preissle state simply that one cannot predict what will happen, that trying to develop and use a template to direct the data analysis is impossible. They feel that "qualitative analysis is interpretive, idiosyncratic and so context dependent as to be infinitely variable. A creative analyst can never be sure that the ending will match the point of view adopted in the beginning" (p.330). In closing, LeCompte and Preissle agree with M&F that a single definition of validity is inappropriate for qualitative research. In a qualitative confirmation process, all authors agree that concerns about validity touch every part of the inquiry. LeCompte and Preissle use triangulation to understand a phenomenon; M&F use triangulation to confirm a hypothesis.

Guba (1981) outlined several paradigms for discovering "truth". These include a judicial paradigm that has well established rules for procedure, rules of evidence and criteria for judging the adequacy of the rationale for a proceeding. This judicial paradigm offers guidelines for behaviour. Another paradigm is that of expert judgement. The third is what he refers to as the rationalistic paradigm and is essentially connected to deductive thinking and a logical positivist point of view. The paradigm he obviously prefers is the naturalistic. The naturalistic is characterized by inductive thinking, and phenomenological views of knowing and understanding social and organizational phenomena. He notes that there are shades of grey in viewing these paradigms and that often they are seen as competing but in the task of knowledge production they are all important. Guba stresses that the naturalistic ecological hypothesis is imbedded in a context which is often more powerful in shaping behaviour than differences among individuals. In conclusion, Guba states that understanding the reality of the world requires acceptance of the notion that the parts cannot be separated. He further concludes that because of the assumptions underlying naturalistic enquiry the traditional concerns for objectivity, validity and reliability have little relevance for the design of the research. The validity of the findings is related to the careful recording and continual verification of the data that the researcher undertakes during the investigative practice. This is consistent with Wolcott

(1990,1994).

The questions of translatability and comparability trouble qualitative researchers (Goetz & LeCompte,1984,1993; Wolcott, 1973). Appropriate sampling improves the generalizability of quantitative studies but researchers improve the quality of their qualitative studies by (among other things) ensuring that units of analysis, concepts generated, population characteristics and settings are described and defined so that they can be compared between studies. Translatability is related to the degree to which the researcher uses theoretical frames, definitions and research techniques that are accessible to or understood by other researchers in similar or related disciplines (Goetz & LeCompte,1984, p228). M&F do not hold contrary views to these discussions of scientific validation; rather they have built upon them in suggesting the possibility of confirmation of findings. Lincoln and Guba (1985), in outlining their fourteen characteristics of operational naturalistic inquiry, write: "naturalistic ontology suggests that realities are wholes and cannot be understood in isolation from their contexts, nor can they be fragmented for separate study of the parts; because of the belief in complex mutual shaping rather than linear causation, which suggests that the phenomenon must be studied in its full scale influence, and because contextual value structures are at least partly determinates of what will be found" (p.39). Spindler and Spindler (1992), when developing their eleven criteria for good ethnography say in Criterion II, "hypotheses emerge in situ as the study continues. Judgments on what may be significant to the study is deferred until the orienting phase of the field study is completed." M&F dodge these issues for most of their book, but they confront the issues in Appendix C.

On the subject of verity and what constitutes rigor, Marshall and Rossman (1989), suggest the use of "controls" which rely heavily on other researchers: the use of a research partner to play devil's advocate; a constant search for negative instances (from Glaser and Strauss, 1967); checking and rechecking the data, and purposeful testing for rival hypotheses; practicing "value free" note taking, and using "strictly objective" observations; devising and applying tests to the data; using the guidance of previous researchers to control for data quality; and conducting an audit of the data collection and analysis strategies (Lincoln and Guba, 1985, p.148). Marshall and Rossman neglect to tell us how a research partner will play "devil's advocate", explain what constitutes "value free" and "strictly objective" observations--nor do they tell us what types of tests to develop, how to develop them, or at what points we should test the data. Negative instances feature in M&F's rules, but their discussion is thin compared to others such as the above.

Talbot (1995), lists a number of factors involved in the *credibility* of findings: remaining in the field over a long period of time; using triangulation; negative case analysis; and having participants review researcher's interpretations and conclusions. She would appear to be in M&F's camp, because she uses the term "confirmability" to describe the process where findings, conclusions and recommendations are supported by the data, and she suggests that there should be an internal agreement between the investigator's interpretations and the actual evidence. Confirmability is only one of four factors that establish *trustworthiness* for Talbot, however, the other factors being credibility, transferability, and dependability. In establishing credibility, she borrows from Goetz and LeCompte (1984) who say that "establishing validity requires determining the extent to which conclusions effectively represent empirical reality, and assessing whether constructs devised by researchers represent or measure the categories of human experience that occur." This definition leans far too heavily on quantitative means in calling for a representation of empirical reality but is explicitly written and comprehensive in description in all other respects. *Credibility* crops up again as Lincoln and Guba (1985), discuss four constructs against which the *trustworthiness* of a study can be evaluated: credibility; transferability; *dependability* and (here's that confirmation again) confirmability. Credibility depends on how accurately the subject is identified and described. Transferability is noted to be impossible from the stance of external validity, but is greatly assisted by providing the greatest possible range of information, and thick descriptive data. The applicability of one set of findings to another setting rests more with the later researcher making the transfer than the original researcher. The authors point out that dependability is difficult to predict in a changing social world. In establishing dependability, the researcher attempts to account for changing conditions in the phenomenon chosen for study as well as changes in the design created by increasingly refined understanding of the setting.

The preface to Strauss & Corbin's (1990) book puts the question of confirmation clearly up front for those who are interested in inductively building theory through qualitative analysis of data. They state that however exciting may be the experience of gathering data, there comes a time when the data must be analyzed and at that time the researcher asks, "How can I have a theoretical interpretation while still grounding it in the empirical reality reflected by my material?" "How can I make sure that my data and interpretations are valid and reliable?" and "How do I pull all of my analysis together to create a concise theoretical formulation of the area under study?" The research question in grounded theory is a statement that identifies the phenomenon to be studied. As the researcher proceeds through the process they can go in different directions and therefore the questions can change. M&F seem able to accommodate this definition of research question and would assure the grounded theory researcher that what needs confirmation is the final conceptualization of the proposed theory and whether or not the symbols of behaviour and language to which the researcher reacted were true. The conceptualization question is one of confirmation; the latter is a question of methodology. Anselm & Strauss strive for *rigor* with seven evaluation criteria embodied in the questions: (1) How was the original sample selected? (2) What major categories emerged? (3) What were the indicators that led to the development of the categories? (4) What categories directed the theoretical sampling process? (5) What were the hypotheses pertaining to conceptual relations among categories and how were these tested? (6) Were there instances in which the hypothesis did not hold up to what was seen (7) How and why was the core category selected and on what grounds were the final analytic decisions made?

We can see within these criteria many of the issues of confirmation and disconfirmation addressed by M&F. They also relate to the criteria for the evaluation of qualitative research proposed by Lincoln and Guba(1985), who were among the first to suggest criteria for good qualitative research. Strauss and Corbin(1990) and Lincoln and Guba (1985) differ in their area of emphasis, with Strauss and Corbin(1990) being more concerned with internal validity and Lincoln and Guba(1985) wanting the work to shed light on other instances. M&F seek to accommodate these different dimensions and to add a strategy that will enhance work that has been started by others.

Return to Contents Table.

Go on to the next section.

Rules for Confirmation

Qualitative confirmation rests on the assumption that there is a hypothesis or at least a hypothetical statement for the study, that data collected through qualitative methodologies will remain qualitative data and that this data can be manipulated within a set of rules. M&F propose just such a set of rules for confirmation of qualitative research. The rules are the application of deductive reasoning in an effort to confirm that the research data logically confirm or disconfirm the hypothesis framed for the study. To apply the rules, therefore, the researcher's first task is to define a hypothesis; it can be either *a priori* or *a posteriori*. The next challenge for the researcher is to determine data collection methods that will produce "evidence-instances". "Evidence-instances" are the result of the qualitative data being recast as evidential statements for confirmation. Here are the rules, exactly as written by M&F (pages 41-42, emphasis in original):

- **Rule 1**: The qualitative evidence instances must be positive for the development of the hypothesis. If they constitute a *denial* of the hypothesis, they disconfirm the hypothesis (see Hempel's 9.3 Df.).
    - 1.1 The limiting "weak" case for confirmation would be the existence of only one positive instance, and for disconfirmation the existence of one (and only one) negative instance.
- **Rule 2**: If the evidence instances constitute a methodologically unique class (8), they must (minimally) not be contradictory to one another.
    - 2.1 As a class, the statements should entail the development of the hypothesis, while the possibility of their own entailment(s) is left open.
    - 2.2 If this class consists of only two instances and they contradict each other, then the hypothesis is neither confirmed or disconfirmed.
    - 2.3 If this class consists of numerous instances, *some* of which are contradictory to the hypothesis, then (2.2) obtains, unless it can be shown that there are *more* instances (positive or negative) and these should be counted for or against the hypothesis, or the above instances should be given a priori "weights" in terms of importance. Note: the assignment of these weights could be given by an agreement of knowledgeable experts.
- **Rule 3**: If relevant background evidence can be adduced for the hypothesis under consideration, and if this evidence alone is sufficient for the development (e.g., entailment) of the hypothesis, and, furthermore, if it is non-contradictory to the class chosen as evidence for confirming the hypothesis, then the given hypothesis may be said to be confirmed.
    - 3.1 If the background evidence is sufficient for the development of the hypothesis but is contradictory to

the evidence chosen for confirmation, the hypothesis' confirmation will remain undetermined.

- **Rule 4**:For a given hypothesis that is to be confirmed by evidence instances derived from a variety of methodological approaches, the class comprising these statements should first be partitioned into relevant categories, i.e., historical-narrative, ethnographic, documentary, quantitative, etc.
  - 4.1 statements within categories should be internally consistent (non-contradictory)
  - 4.2 Each partitioned subset need not be sufficient for the derivation of the hypothesis, but the totality of subsets comprising the class should be sufficient (and, hopefully necessary).
  - 4.3 Depending on the number of subsets, if one subset contradicts the others, either partially or wholly, the confirmation of the hypothesis is left undetermined.
  - 4.4 If one subset is disconfirming to the hypothesis but neutral or non-contradictory tothe remaining subsets, and the remaining ones are consistent, the hypothesis will be considered confirmed.
  - 4.5 The hypothesis will be considered disconfirmed (i.e., rejected) if (a) all subsets disconfirm, (b) if a majority disconfirm, or (c) in the limiting case of two subsets, where one is contradictory to the other, if additional grounds (i.e., agreement by experts) can be adduced for the disconfirming subset, this will be taken as sufficient for disconfirmation. (In this case the evidence has been "weighted" towards disconfirmation; of course, the converse, i.e. for confirmation, could be similarly argued.)

Return to Contents Table.

Go on to the next section (Application of Rules).

Application of the M&F framework to the Sandra Rubin Glass study, "Markets and Myths: Autonomy in Public and Private Schools", published in *Education Policy Analysis Archives*, vol. 5, no. 1, January 6,1997 (the 'Glass' study)

The title describes the article well: Glass undertook a large-scale study of teacher and principal autonomy in public and private secondary schools and used the data to explore claims made by Chubb and Moe (1990). She challenges the latter's assertions: that the organization of private schools offered greater teacher autonomy resulting in higher student achievement and that the bureaucracy of public schools stifles autonomy and limits student achievement. Glass attempted to bring to the surface conditions which constrain teacher and principal autonomy in both public and private schools. Although she did not express it this way, it is reasonable to state that she had a hypothesis: "There are no differences between public and private schools in the amount of autonomy teachers and principals have." She also set out to answer the question, "What conditions impede teacher and principal autonomy in both public and private schools?", implying the hypothesis: "There are no differences between public and private schools in the conditions that foster or inhibit autonomy."

In her comprehensive study, Glass used both data source and methodological triangulation. "The methods employed in this investigation were those of the multi-site qualitative case study: interviews from multiple data sources, observations and field notes from a variety of on-site meetings and visits, and analysis of documents (brochures, teacher handbooks, policy manuals, meeting agendas)." She conducted an intensive study of three public and three private secondary schools, interviewing fourteen private school teachers, fifteen public school teachers; assorted principals, heads and assistants from each school were interviewed at their respective sites.

According to M&F, triangulation is described as a series of strategies that directs both the generation of data and the clarification of findings . They go on to say that the purpose of triangulation in qualitative inquiry is "to provide a rationale for increasing the plausibility of qualitative findings" (M&F, p.28). M&F state that, "if different data sources are used for the study of a particular problem, and if it can be claimed that they are relevant for the problem, the likelihood for the total data set to reflect reliability and validity will be enhanced" (p.27). The different sources listed by Glass have face relevance to the problem and in principle enhance the likelihood of reliability and valididy. In the journal article, however, the only data we can identify by source is from the interviews. A set of hyperlinks allows readers access to them as the findings are being reported. Whenever a quotation is given from an interview, the reader may choose to examine the quoted passage in its full context by clicking with the mouse on an icon at the margin. (Glass has taken the extra step of directing the link to the exact location of each quotation in the interview and highlighting the quotation in bold.) We thus have explicit examples of the "evidence-instances" referred to in M&F's rules and an unprecedented opportunity to assess in detail whether the evidence is "positive for the development of the hypothesis". The thirty-seven interviews are presented in a standardized pattern, clear and accommodating to read. She used many of the interview questions from the Moles (1988) and Blase's (1991) surveys (part of the 'High School and Beyond' survey) used by Chubb and Moe (1991) to develop their index of teacher and administrator autonomy. The use of structured tools that have been previously tested lends credibility to the data set. Participants were selected based on years of teaching experience (at least five), and years of experience at the present school (at least three). The school sites were chosen so that the constituencies of both the public and private schools were as comparable as possible; in both private and public situations, schools servicing high income families and focusing on academic excellence and college preparation were selected as well as less favoured districts. Glass does not explicitly state how the data will be used in the confirmation process (she did the work before M&F's book was published), but she does share her process of weighing the evidence. For M&F, this consideration is a necessary step in research design that allows for consequent evaluation in relation to qualitative confirmation.

We now assess the Glass study for qualitative confirmation by using the M&F rules.

As noted in the previous section, **Rule 1** states that the evidence instances must be positive for the development of the hypothesis. If they constitute a denial of the hypothesis, they disconfirm the hypothesis. In more than thirty interviews with teachers and their principals, from both private and public schools, Glass found that participants from both private and public schools experienced about the same measure of autonomy in their environments or were able to work around conditions that constrained it. The interviews brought out the degree of complexity inherent in the idea of autonomy. Rule 1 is thus satisfied.

Rule 2 states that if the evidence instances constitute a methodologically unique class, they must (minimally) not be contradictory to one another. The interviews constitute just such a class, and the evidence instances Glass presents do not contradict each other. In most journal articles, or even books, we have to be content with the evidence instances provided by the author-always and necessarily a small subset of all possible instances-but here we are able to read all the text of all the interviews. We do not claim to have done so, but each one of us read at least one interview through, explicitly searching for evidence instances that would contradict the hypotheses. We found none. While the researcher does not report explicitly on the data realized from on-site meetings, visits, and analysis of documents, she does refer to the high achievement standards of schools, curriculum content, parent involvement, and how arrangements were made to collect data--information that had to be collected from these sources. Consider, for example, the following statements from Glass's "Findings" (we have put inferences assumed to be made from field notes and the like in bold):

In a private school, new teachers will generally define the curriculum predicated on their own content knowledge and interest. Because of smaller faculty numbers, there may be two or three other teachers with whom to coordinate curriculum; yet each teacher specializes in a particular facet of that content area. While each of the three independent schools in this study have either a middle school or middle and elementary school as part of its organization, **students come from a variety of other schools**. Consequently, **coordination is a matter of interest only within the upper school**. Any coordination of curriculum is accomplished within the institution, as described by this private school teacher: (a quotation follows, with link to the interview).

....

This study was conducted in a right-to-work state in which teacher unions are virtually non-existent, **but teacher associations are predominant**. These **associations are seen as variously strong or weak depending on locale**. Only **one of the three public schools is in a district having a very strong teacher association. Most, if not all, of its teachers are members of the association and quite a few are active in its leadership. The other two schools are in districts that negotiate teachers' contracts with the association, although the faculty are much less active.**

It is not possible to confirm or disconfirm these statements from the journal article, because the relevant evidence instances are not supplied. Those that are supplied satisfy Rule 2, but what do we say about the others? We do not know.

The Chubb and Moe (1990) findings are presented, and we could regard them as "background evidence" (**Rule 3**), but rather than being "adduced for the hypothesis", they are findings to be questioned, and possibly to be disconfirmed. This is an outcome of research that M&F appear not to have foreseen. Glass's opening statement, and her analysis of data is an argument for a disconfirmation of their report. On the other hand, background information from Sedlak (1986) and Ball (1987) are adduced as support for Glass's hypothesis. In our opinion, the evidence of

Sedlak and Ball is not sufficient for development of the hypothesis, but neither are they contradictory. Applying Rule 3, we would say the hypothesis is confirmed by Sedlack and Ball but contradicted by Chubb and Moe. The hypothesis would therefore remain as "undetermined" if the Chubb and Moe data were sufficient. This is why Glass argues so strongly that the Chubb and Moe evidence is, to put it mildly, not sufficient. She points out that they present no evidence whatsoever on private schools from the "High School and Beyond" study that is the basis for their arguments. Applying Rule 3, we conclude that the background evidence is itself contradictory and would leave the hypothesis undetermined. It seems as if we have to decide for ourselves how to weight the evidence.

Rule 4 states that, "for a given hypothesis to be confirmed by evidence instances derived from a variety of methodological approaches, the class comprising these statements first should be partitioned into relevant categories" (p.43). In the Glass study, as previously stated, the interview data are internally consistent, and we do not have access to the data from the field notes, on-site meetings and visits or document analysis. Rule 4, therefore, cannot be applied. Glass has woven information in her discussion that makes it reasonable to acknowledge consistency in those data, but the evidence instances are missing that would allow us to make a strong case.

We believe that the Glass study, in employing focused interviews using open ended questions and observations, makes a plausible case for qualitative confirmation of her hypothesis, but applying the M&F rules did not firmly settle the matter. The study explores real life situations with no attempt to manipulate or control conditions. She argues that a high degree of autonomy is experienced by teachers and principals in both private and public schools, and that her findings disconfirm those of Chubb and Moe (1990) that teachers in private schools experience more autonomy than teachers in public schools. Regarding her second hypothesis, she identified six factors associated with autonomy: conflicting and contradictory demands, shared beliefs, layers of protection, a system of laws, funding constraints, and matters of the size of institutions. She concludes that autonomy is a complex process--an issue that does not distinguish the public from the private sector. As her only qualifier, Glass observes that similar organizational effects may not be encountered in schools under the duress of poverty and social dislocation, perhaps seeking to avoid the fallacy of confirming the consequent. It would appear from this example that qualitative confirmation yet eludes us; we revisit this in our appreciation.

Return to the Contents Table.

Go on to the next section (Appreciation).

An appreciation of the book by Miller and Fredericks, *Qualitative Research Methods*

Some of us attended graduate school and had our early experience in an educational research that was dominated by experimental psychology. Prompted by Campbell and Stanley (1963) we were concerned with the validity of our research, but we believed that if only we found the correct experimental design and carried it out competently our results would be valid. With methods based on the probability calculus, we could arrive confidently at a "significance level." Others of us know little of statistics and experimental design, having studied in programs and with faculty who do not believe in realism and seek, for example, verstehen-"a type of historical or contemporary insight which cognitively reconstructs a plausible interpretation of an action or event given knowledge of the cultural 'rules'." (M&F, p. 74) It may well be that the latter group is now dominant in educational research, and many are quite confident in their methods and believe their results adequately justified. Some believe no more justification is possible (e.g., Smith and Heshusius, 1986). At least a small subset of researchers remains uneasy that qualitative approaches lack means for validation. It is this group that M&F addresses, of course, those that would be extremely happy to have a means to attain "qualitative confirmation." In our opinion, graduate students (especially those considering qualitative methods for their research) should study and assess M&F's rules and rationale, whatever camp they are in. M&F's presentation makes this much more difficult than necessary, but the issue is important enough to make the effort.

The overall organization of chapters is logical; once apprehended it is clear:

1. The major issue in qualitative inquiry (confirmation-and introduction to rules)
2. Hypotheses in qualitative research methods (defense of hypotheses and the nature of evidence)
3. Additional rules of confirmation (M&F's version plus discussion of disconfirmation)
4. Assessing qualitative studies (applying the rules to some published reports)
5. Epistemological asides and conclusions (revisiting the intellectual roots)

Reading through the chapters, however, one finds complex concepts and intricate arguments that can only be clarified (usually!) by reading the appendixes:

1. Some notes on the nature of indeterminacy
2. Clarifying the "adequate evidence condition" in educational

theory and research
3. Reciprocal paradigm shifts and educational research: A further view of the quantitative-qualitative dilemma.

After all this work, however, you will find that qualitative confirmation is still a judgment call. We attempted to illustrate this with our own application of the rules to the "Glass study". The published report of that study was attractive to us because the "evidence instances" were so obvious and because they constituted a "methodologically unique class". Because all the data (interviews) was available, we could do some verification not usually available (search for negative instances not reported in the published version, for example). The studies analyzed by M&F in Chapter 4 did not provide as good a test of the method, but even this good test ended inconclusively, IOHO.

As with all works of genuine scholarship, one of the benefits of study is the acquaintance (or reacquaintance) it provides with key scholars and their ideas. M&F write from outside the "college" known best by the authors of this review, and we found many new and insightful sources. Most of us have heard of Quine but have not studied his books. Our inadequate preparation in philosophy left us ignorant of the seminal contributions of Winch, and we agree with M&F that, "It is amazing how much debate has been generated by the two rather modest works of Winch (1958, 1964)." (p. 74) Swinburne lurked unread in our library, as did Roth (shame, shame). We were led to an even stronger appreciation of Wolcott and introduced to the "cognitivist and semanticist", George Lakoff.

In summary, the presentation is sloppy, the writing dense and the organization suboptimal, but the topic is important and the fresh perspective welcome. We wish the authors would bring out an new and improved edition, but even if not we recommend the book to you.

Return to the Contents Table

References

Ball, S. J. (1987). *The micro-politics of the school.* New York, NY: Methuen and Company.

Blase, J. J. (1991). Some negative effects of principals' control-oriented and protective political behavior. *American Educational Research Journal, 27*, 727-753.

Campbell, D. & Stanley, J. (1963) Experimental and quasi-experimental designs for research on teaching. In, N. Gage (Ed.) *Handbook of research on teaching*. Chicago: McGraw-Hill.

Carnap, R. (1962). (2nd ed.). *Logical foundations of probability*. Chicago: University of Chicago Press.

Chubb, J. E., & Moe, T. M. (1990). *Politics, markets, and America's schools*. Washington, DC: The Brookings Institution.

Collin, F. (1985). *Theory and understanding: A critique of interpretive social science.* London: Basil Blackwell.

Coombs, C. (1964) *A theory of data*. New York: Wiley.

Davidson, D., (1984). *Inquiries into truth and interpretation.* Oxford: Oxford University Press.

Denzin, N.K. (1978). *The research act*. New York: McGraw - Hill.

Dreher, M. (1994). *Qualitative research methods from the reviewer's perspective, Critical Issues in Qualitative Research Methods.* J. Morse (Ed.), (pp.281-299). London: Sage.

Fuller, S. (1988) *Social epistemology*. Bloomington, IN: Indiana University Press.

Fogelin, R. (1982). (2nd ed.). *Understanding arguments*. Toronto: Harcourt Brace Jovanovich Inc.

Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory*. Chicago: Aldine.

Goetz J., & LeCompte, M. (1984). *Ethnography and qualitative design in educational research*. Orlando, Fl: Academic Press.

Greenwood, J. D. (1989) *Explanations and Experiment in Social Psychological Science: Realism and the Social Constitution of Action*. London: Springer-Verlag.

Guba, E.G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Communication and Technology Journal, 29*, 75-91, as cited in Owens, (1982). Methodological perspective: Methodological rigor in naturalistic inquiry: Some issues and answers. *Educational Administrative Quarterly, 18*(2),1-21.

Hempel C. (1965). *Aspects of scientific explanation and other essays in the philosophy of science.* New York: Collier - MacMillan Limited.

Lakoff, G. (1987) *Women, fire, and dangerous things: What categories reveal about the mind.* Chicago: University of Chicago Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by.* Chicago: University of Chicago Press.

LeCompte, M.D., & Preissle, J. (1993). (2nd ed.). *Ethnography and qualitative design in educational research.* San Diego: Academic Press.

Lincoln, Y., & Guba, E. (1985). *Naturalistic inquiry.* Newbury Park: Sage.

Marshall C., & Rossman, G. (1989). *Designing qualitative research.* Newbury Park: Sage

Miles M.B., & Huberman, A.M. (1984). Driving valid meaning from qualitative data: Toward a shared craft. *Educational Researcher, 13*(5), May 1984, 20-30.

Miles, M.B., & Huberman, A.M. (1984). *Qualitative data analysis.* Beverly Hills: Sage.

Miller, Stephen I. and Fredericks, Marcel (1994) *Qualitative Research Methods: Social Epistemology and Practical Inquiry.* New York: Peter Lang.

Moles, O. (Ed.) (1988). *High school and beyond administrator and teacher survey (1984): Data file users manual.* Washington, D.C.: Office of Educational Research and Improvement, U. S. Department of Education.

Owens, R.G. (1982). Methodological rigor in naturalistic inquiry: Some issues and answers. *Educational Administrative Quarterly, 18*(2), 1-21.

Polit, D. & B. Hungler (1995). *Nursing research principles and methods*, Philadelphia: J.B. Lippincott.

Popper, K. (1969). *The logic of scientific discovery*. New York: Science Education.

Quine, W. (1960). *Word and object*. Cambridge: Massachusetts Institute of Technology.

Quine, W. (1963). (2nd ed.). *From a logical point of view*. New York: Harper books.

Rosch, E. (1975). Cognitive representation of semantic categories. *Journal of Experimental Psychology: General, 104, 192-233.*

Roth, P., (1987) *Meaning and method in the social sciences: A case for methodological pluralism*. Itaca, NY: Cornell University Press.

Rosenberg, A. (1988). *Philosophy of social science*. Boulder, CO: Westview.

Scheffler, I. (1965) *Conditions of knowledge: An introduction to epistemology and education*. Glenview, IL: Scott, Foresman and Co.

Sedlak, M. W., Wheeler, C. W., Pullin, D. C. & Cusick, P. A. (1986). *Selling student short: Classroom bargains and academic reform in the American high school.* New York, NY: Teachers College Press.

Smith, J., & Heshusius, L. (1986) Closing down the conversation: The end of the quantitative-qualitative debate among educational inquirers. *Educational Researcher, 15*, 4-12.

Spindler G., & Spindler, L. (1992). Cultural process and ethnography: An anthropological perspective. In M. LeCompte, W. Millroy, & J. Pressile (Eds.). *The Handbook of Qualitative Research in Education*. San Diego: Academic Press.

Stake, Robert E. (1995) *The Art of Case Study Research.* Thousand Oaks: Sage.

Strauss, A., & Corbin J. (1990*). Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park: Sage.

Swinburne, R. (1973). *An introduction to confirmation theory.* London: Methuen.

Talbot, L. (1995). *Principles and practices of nursing research.* St. Louis: Mosby.

Winch, P. (1964) Understanding a primitive society. *American*

Philosophical Quarterly, 1, 307-324.

Winch, P. (1958) *The Idea of a Social Science and its Relation to Philosophy*. London: Routledge and Kegan Paul.

Wolcott, H.F. (1973*). The man in the principal's office: An ethnography*. New York: Holt, Rinehart and Winston.

Wolcott, H.F. (1990). *Writing up qualitative research*. London: Sage.

Wolcott, H.F. (1994). *Transforming qualitative data*. London: SAGE. Especially Chapter 11, On seeking-and rejecting validity in qualitative research, 337-373.

Return to:

- Purpose
- Description
- Related Works
- Rules
- Application
- Appreciation,
- Contents

Notes

(1) This is a collaborative review, so the authors are listed alphabetically.

Les McLean
lmclean@oise.utoronto.ca
Professor, Department of Curriculum, Teaching and Learning
Ontario Institute for Studies in Education/University of Toronto:
OISE/UT
(416) 923-6641, ext. 2478

Margaret Myers
lmmyers@gtn.net
Doctoral Student, Department of Curriculum, Teaching and Learning
OISE/UT

Carol Smillie
csmillie@oise.utoronto.ca
Doctoral Student, Department of Curriculum, Teaching and Learning
OISE/UT

Dale Vaillancourt
dvaillancourt@oise.utoronto.ca
Doctoral Student, Department of Curriculum, Teaching and Learning
OISE/UT

(2)Quine uses the term "translation" in the usual sense of finding a representation in a language of a text expressed in another, broadening and deepening the discussion in *Word and Object* (1960). The term "translation manual" has been extended to describe how a researcher understands what people in a "foreign" culture say and do, how we make sense of field notes, for example. There will always be more than one possible translation manual for any situation, just as there is more than one possible translation of a text from English to French. That said, we may argue in favour of one particular translation manual, just as we may say we prefer one translation over another.

(3) "We will also use the phrase 'evidence-instance' to indicate that the qualitative data are now being recast as evidential statements for confirmation." *M&F*, p. 41.

(4) Recollected narrative-call it "personal communication".

(5) "obervation" (p. 8). "inherit" instead of "inherent" (p. 16). "accept" where "except" is intended (p. 22). "was" when "were" is correct (p. 44). view**ed** (p. 45). "this not entail ..." does need

"does" (p. 94). "or" (not 'of') (p. 119). "intact" rather than "in tact" (p. 145). References are missing from the end of Chapter 3, e.g. Miller (1990)!

(6) M&F argue in the same spirit as Clyde Coombs (1964), who distinguished between "observations" and "data" (he dealt only with observations in the form of numbers). What M&F call data, Coombs called observations, which only became data after application of one of the scaling techniques coming to fruition and/or being developed by Coombs. The spirit is that information, whether qualitative or quantitative, comes to every researcher first in a raw form that must be refined before it can be used to make inferences.

(7) Some of Hempel's rules (from M&F, p. 11):

9.1 Df. An observation report B directly confirms a hypothesis H if B entails the development of H for the class of those objects which are mentioned in B.

9.2 Df. An observation report B confirms a hypothesis H if H is entailed by a class of sentences each of which is directly confirmed by B.

9.3 Df. An observation report B confirms a hypothesis H if it confirms a denial of H.

9.4 Df. An observation report B is neutral with respect to a hypothesis H if B neither confirms nor disconfirms H.

(8) By a 'methodologically unique class', we mean a situation where the researcher employs one dominant form of data collection, such as interviews, for instance. While such a situation is probably not realistic, it is a logical possibility and for this reason is included. (p. 42).

Return to:

Purpose

Description

Related Works

Rules

Application of Rules

Appreciation

EPAA
• enter the archives • browse the abstracts • the editors • the edit board
• submit article • submit commentary • search • subscribe
volume: • 1 • 2 • 3 • 4 • 5 • 6 • 7

This article has been retrieved 1108 times since June 13, 1997.

Education Policy Analysis Archives

Volume 5 Number 14 June 13, 1997 ISSN 1068-2341

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass Glass@ASU.EDU.

College of Education

Arizona State University,Tempe AZ 85287-2411

Copyright 1997, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

Rhetoric and Educational Policies on the Use of History for Citizenship Education in England from 1880-1990

Jessie Y. Y. Wong
Nanyang Technological University

Abstract

This article attempts to review the rhetoric and the educational policies on the use of history for citizenship education from 1880-1990 in England. In many instances, the rhetoric served as powerful tools to gain the support of educational authorities, namely, the Board of Education, Ministry of Education and Examination Boards. Their support was reflected in the change of educational policies and school syllabi that followed. This study shows that there was strong and consistent widespread rhetoric on history's contribution to citizenship education throughout the century, neither stopped by the two great wars nor impeded by the challenge of social studies as a citizenship subject after the Second World War. Instead it was challenged by the discipline itself in the early 1980s when some historians began to doubt the "new" history on the ground that the "real" history was being devalued. Consequently, there was evidence that the "new" history did not take off widely. In many schools, history was taught for its own sake. Its value for the education of modern citizenship was not being emphasised. This article ends with the argument that under the environment of the National Curriculum, first implemented in the country in 1989, history still claims its relevance for citizenship education.

History for Moral and Patriotic Citizenship Education

Claims for history as a citizenship subject dated back long before it became an academic discipline. When Herodus, the father of history began to write history, it was for its social relevance. Since then, history has experienced a number of phases of development but its utility as a citizenship subject continued. Early this century, F. J. Gould, a British Educationist and humanist, published a large quantity of literature using history as a source for citizenship education --- promoting moral and patriotic citizenship in young children. Other publications from historians and from the Board of Education also demonstrated the value of history for moral and patriotic citizenship education.

Gould, different from a modern historian, saw history as an all-embracing subject which could, properly taught, provide social, moral and civic education for training children to be good future citizens. His approach was not only patriotic but also humanistic. He proposed seriously the use history for the training of patriotism and social loyalty. At the same time, he was aware of the importance of an international perspective. Through story-telling and through history, he worked out numerous schemes for the training of citizenship. He provided more than just rhetorical support for the subject. It was evident in the 1904 Elementary Code that history was approved as a school subject because of the emphasis it gave to the lives of great men and women, and the lessons to be learned from them, by means of which the characteristics of the good citizen were thought to be inculcated. Gould's work indicated the potential he saw of history's capacity to promote citizenship education. His writings began to be published for this purpose from the 1880s until he died in 1938.

In fact, when history was first introduced in the school, its only objective was, like what Gould had often mentioned, to inculcate those values, social and political which the nation and the people had come to accept as characteristic of itself. As time passed, history was seen futile as a school subject except in providing citizenship education. Examples could be found in Laurie's work in the late nineteenth century. Laurie (1867, p.144) wrote:

> "To the school-boy ... [History] is of value in so far as it brings to his knowledge wonderful deeds done in the discharge of patriotism and duty. In all other respects, it is utterly barren of good results, and involves a futile expenditure of valuable school-time. A dim outline of royal genealogies, of dates, the intervals between which are full of plottings and counter-plottings, and of parts which, however capable of interpretation by the matured capacity ... to the raw experience of the child or the boy, little more than an exhibition of the worst passions that afflict humanity, and all these epitomized into small compass and only partially and fragmentarily acquired --- such is school history. It seems to us, therefore, that the reading of history in the primary school is little better than an abuse of time."

Of course, Laurie had a narrow view of citizenship education. He believed that the training of patriotic, dutiful and passive citizens was the only goal of learning history. His *History in the Primary School* was published nearly four decades earlier than most of Gould's work on history and training for citizenship. His work showed that history was the subject for citizenship education before Gould. History could served this purpose by mere dispensation of knowledge. It was true that school history has all the time included the acquisition of a body of knowledge: facts of events, dates, great names, etc. This knowledge, filled with false representation, bias and imperialistic feelings, was all that was thought necessary to provide models for future citizens.

By the last decade of nineteenth century, some historians and educationists had begun to be more alert as to a wider scope of history for citizenship education, to lead away from partiality, bias and the false representation of the patriotic approach. Pitt published *English History, with its Wars* left out in 1893. It was a reflection of a change of attitude towards history for citizenship education, believing that 'drum and trumpet' history could no longer achieve the objective of training for future citizens. Pitt's book marked the beginning of history textbook revision.

Consequently, in the twentieth century, history had increased in popularity both in the universities and in the schools because of its change of content and its new role to disseminate a wider perspective of citizenship education. In particular, by the time of the 1902 Act, the teaching of history within the curriculum of both public elementary and secondary schools had begun to lay particular emphasis upon political and constitutional aspects which led many historians to believe that the teaching of civics was the special task of history teachers. The primary aim behind the framework of many history textbooks began to change. Most would emphasize the

> "inculcation of the ideas of citizenship which themselves developed from a mere prescription of various rights and duties to that of equipping the future citizen with a full knowledge of how his society had evolved." (Cook, 1978, p.42)

Having to know how society had evolved required no short history. The consequence was the encouragement of the discernment of 'movements' and 'trends' over large sweeps of history in the school syllabi in order to convey an awareness of human progress and to foster a realization of the past as an essential prelude to the present. It was generally urged that the pupils should be led to a greater awareness of the national society in which they lived and how it had developed to its present stage. Thus the emphasis on national citizenship was clearly shown. For example, Withers (1904, p.200) wrote:

> "It is because of its bearing on the future of our civic and national life, even more than on account of its value to the imagination and the understanding that the study of history may claim an honored place in the timetable of our primary schools."

It was in 1905, when Bourne's famous book, *The Teaching of History and Civics in the Elementary and the Secondary School* was published, that history's role for civic and citizenship education was re-emphasized. Bourne was the well-known professor of the College for Women at Western Reserve University of the United States. Most people involved in education at the time, whether in England or North America would have accepted fully or in part his contention that civic and citizenship education was the responsibility of history and that

> "pupils may be instructed in the duties of citizenship in two ways: First, by studying the structure of government and the duties of the individual in relation to it (that is to be taught through history lesson), and second, by discipline in the performance of such social duties as fall to them during school life, with the expectation that thereby sound habits may be created and good citizenship may be only a continuation of the earlier training in conduct." (Bourne, 1905, pp. 93-94)

In all cases, history textbooks tended to merely describe institutions and offices of government. Bourne's book was different. He called for a more active involvement of pupils within and without the school to make training for citizenship more effective. He was very much in advance of time. His work had not only influenced the Americans but also the British in looking at history's contribution to citizenship education. In Britain, evidence of a new dynamism in historical studies was indicated, at school level, by the Board of Education's publications. The first edition of *The Handbook of Suggestions for Teachers* in 1905 included a chapter on the problems, objectives, content, methods and aids of history teaching in elementary schools. This chapter offered a fresh look at the new ways of history teaching under which its contribution for the training of future citizens was reconsidered in favor of Bourne's suggestion. For secondary schools, the introduction of new ideas along these lines was also evident. The Board's circular, *History*, published in 1908 and *Modern European History* published in 1914, both offered some cautiously progressive ideas in content, notably in local and European history and in methods.

The catalyst of, and the principal pressure group for, this new-found ferment in the teaching and study of history was the Historical Association, founded in 1906 by a number of distinguished teachers in schools and universities. Thus in 1909, this Association indicated its commitment to an important role for history in citizenship education by publishing the leaflet No. 15 entitled, The Teaching of Civics in Public Schools.

The notion of nationalism was the most important part of citizenship education then. Thus in the Wiltshire Memorandum 1910, it stated that:

> "The amount of information a child can retain is small, and efforts to make it extensive are thrown away, but it is most important that a child should realize before he leaves school how his country is governed, what he inherits from the past, and what duties he owes as a patriot, and a citizen". (1910, pp. 16-17)

The Wiltshire committee realized the massive amount of facts in history. Therefore the selection of content on agreed criteria was the only sensible way to reduce it. The general consensus seemed to be that anything could go except national history. Thus the emphasis on history's role in citizenship education remained at the national level of citizenship.

Later, Hayward, a supporter of public elementary schools offered a more progressive thought in citizenship education. He noted that mere dispensation of historical knowledge was an insufficient and ineffective way of training for citizenship. Thus he urged for more stress to be laid on the transmission of values such as

> "the high sense of duty, the patriotic devotion, the subordination of self-interest to the good of the community and the magnificent conception of public service." (Hayward, 1910, p. 355)

Very much involved with the great public schools, he promoted the transfer of the spirit of great public schools to the elementary schools. He acknowledged that the moral climate of the school was as important as historical facts in promoting good citizenship. This idea of his is still present in the minds of many modern educationists. The ethos of the school is still believed to be an important but hidden aspect of citizenship education today.

After the First World War, history continued to play an important role in citizenship education. However, there appeared a number of publications which geared towards a more positive form of citizenship education. The war had definitely had an impact on the type of citizenship education which history should offer. Hughes's *Citizen to Be* was a typical one, in which she expressed the hope that the elementary history syllabus should rid itself of "the details of remote wars, of court intrigues, of royal pedigrees, of much constitutional history". Hughes' book was meant to be used in the teaching of history. Her shift from viewing history for promoting patriotic citizenship to more progressive citizenship was characteristic of the beginning of a general shift away from more chauvinistic views in citizenship education after the War; thus marking the beginning of a new notion to teach progressive and international citizenship in history. The following decade saw the use of history to promote international understanding. The purpose was to avoid conflict and to prevent another world war.

History for International Citizenship

Significant changes in the attitude towards using history as a form of citizenship education were necessitated by post-World War I society. An history which reflected nationalism, and which was capable of producing law-obeying citizens of the nation was seen as insufficient to lead the new generation into social reconstruction. On the one hand, there was the pressure to introduce an expanded concept of citizenship education. On the other, there were variations among the historians as to how history could best be used to transmit citizenship education.

During this difficult time, two polarized ideas about history teaching for citizenship could be traced. Firstly there was the 'horrors of war' --- the pacifist school of thought, which believed that history teaching could help to avoid war, and argued that nationalism was the prime cause of war. So narrow, nationalistic history content should be removed from the syllabus. Protagonists of this school of thought believed in the ability of history to promote international citizenship and thus to avoid war. They advocated the study of world history which would foster international cooperation and social progress. Secondly, there existed a continuing support for the study of national history --- the nationalist school of thought. It stressed patriotic pride in the navy, the army and the unity of the empire. It accepted history's contribution to the education for imperial citizenship and showed little enthusiasm for change. This group had probably unconsciously launched an attack on the pacifist school of thought in history teaching as early as 1916 at the A.G.M. of the Historical Association, where all speakers favored the teaching of naval history, and the whole audience was 'exhilarated with the imperial ideas'. The only organization at that time which was in support of the pacifist school of thought was the League of Nations Union which argued for the inclusion of the aims and instruction of the League of Nations in history. History syllabi according to this union should be purged of war if they were to become an effective instrument for peace.

This nationalist school of thought was as emphatic in preserving imperial ideas as the pacifist school was in promoting change. However, both schools had some common grounds of agreement on history teaching, and on history as an instrument for citizenship education. They both agreed that the past could be used to explain the present, therefore history teaching must include contemporary social, economic and political issues, not least citizenship issues.

The belief in the efficacy of history to explain and draw conclusions spread far beyond the university teaching of the subject. In 1918, despite the difference of ideas between the pacifist group and the nationalist group, the discipline of history entered a new period of popularity. It stood in high esteem in universities and schools and with the Board of Education, examination boards and the general public. With the expansion of the syllabus to include contemporary European and world history, and often that of each school's locality, a new relevance was promised. History could then be used to foster local, national, European and world citizenship. These underlying objectives for history were reinforced in 1923 by the Board of Education, in its Pamphlet No. 37, The Teaching of History, which stated that the period after the First World War was a period

> "in which the arguments for studying history, both on the civic and international side, have been brought home to us and intensified in a way never possible before." (HMSO, 1923, p. 9)

This was a period of expansion of the concept of citizenship whether taught in history or in geography. This history pamphlet went on to give a clear view of the character of history teaching which the primary and secondary schools should follow. In the primary schools, pupils should follow a course of general stories drawn from all countries, by which the first interest in history should be inspired. In the secondary schools, an unbroken course in English history from the first form onward to the first school examination was proposed. The work should be planned in line with the Board's Circular on The Teaching of History in Secondary Schools, No. 599, 1908. Thus advocated the teaching of everything in English history, with European history as ancillary to it. The advance in history in the schools had also had, to some extent kept pace with the public desire to study the history of other countries which had been stimulated by the First World War. The aim of history teaching, therefore, was to implant firmly in pupils' minds what the committee considered as fundamental outline and to extend their range of ideas, enrich their minds and give them a world-wide interest. In modern terms, it would be to foster international citizenship. Accordingly, in the advanced course, especially at "A" level, a period of English history and a corresponding period of European history should be given. In this way, it should provide history a wider opportunity to cover the expanded notion of citizenship of the inter-war period.

Equally, history teaching was intended to have concern for social and civic issues. The pamphlet devoted a section to social and civic education in which it discussed the more recent development in history for this purpose. Social history was given attention in all classes. It was hoped that through the study of the account of the developing life of the whole people of the country, some kind of social education could be promoted. This connected well with the increased attention given to local history which was another growing feature in history at that time.

The concern for civic and social issues in the teaching of history indicated that the teaching of civics as a separate subject was not taken seriously. Most teachers, unlike those in the United States, preferred to allow the lessons which civics would impart to flow naturally from the ordinary school history course. The general opinion was that history should contribute to civic education. Thus it followed that some teaching of modern history and modern problems was vital in any history course.

Throughout the 1920s, optimism for history both as a subject in itself and as a citizenship subject continued. It was argued by Hadow committee in 1926 that the study of history was of "first importance". This positive view reached a peak by the end of 1920s and continued to be strong in the 1930s. The protagonists of citizenship education (The League of Nations Union, The Association for Education in Citizenship) did not hesitate to seek more support from all subjects to promote the study of citizenship in schools. From the two major books published by the Association in the 1930s, it was clear that the Association sought to promote citizenship education through all subjects in the school curriculum.The League of Nations Union in Britain also continued their campaign to encourage the subject lobbies to include the aims and instruction of the League of Nations in their teaching, as a device for promoting, in particular, international citizenship. History was no exception to include the Covenant of the League in its instruction to teach citizenship education.

The teachers were urged to take care to use the words and phrases found in the covenant with a certain frequency in general instruction, and thus informally prepared the way for their employment in the recital of the League's activities. The ultimate goal was for them to assimilate and practice these values. Most of them were related concepts of international citizenship, such as cooperation, peace, security, obligation, law, justice, sincerity, representations, voting, aggression, political independence, arbitration, convention, decision, dispute, mutuality, freedom of conscience, mitigation of suffering, etc.

The League advocated history as a subject capable of dealing with international citizenship issues. In its agenda for the second session in 1935, the Advisory Committee on League of Nations Teaching drew up a program for the teaching of history. Changes in the conventional history teaching were made necessary. It also endeavored to teach international relations and the League of Nations. Its interest in using history to cover citizenship education indicated the preconceived notion of history as a citizenship subject. In the following year, a report was made on The League of Nations and Teaching of History and Geography. In it, suggestions were given on how history and geography could be used as a tool for teaching the League's notion of citizenship. In Britain, the League of Nations Union was in the position to carry out promotional work and provide help to teachers. It committed that history

> "undoubtedly offers ample opportunity for implanting in the hearts and minds of young people League ideas, such as the guaranteeing of an international system of law involving the limitation of national sovereignty in consequence of the renunciation of war as a final argument, the necessity for collaboration between peoples in the economic and intellectual sphere, and the ideal of humanity and peace." (League of Nations,1936, p. 16)

The Association for Education in Citizenship, established in

UK in 1935, was another organization which added pressure to introduce citizenship education into the school curriculum. Having failed to set up a separate subject for citizenship education, its alternative method was to give impetus to other subjects to cover this area. The Association published *Education for Citizenship in Secondary Schools* in 1935. In it Doyle claimed history's responsibility for the education of a democratic citizen. Four years later, it published *Education for Citizenship in Elementary Schools*. In it, Strong made a similar statement:

> "If it comes as some surprise to certain readers of this book that there are so many subjects in the elementary school curriculum through which it is possible for citizenship to be taught, there are, on the other hand, probably few teachers who would be disposed to deny that history is the most positively civic subject of all." (Strong, 1939, p. 112)

When the Second World War broke out in 1939, it was assumed that neither education for international understanding nor education for democracy sufficed in itself as an effective mode of citizenship education, judged as peace education. Neither the League of Nations nor the Association for Education in Citizenship succeeded in their common aim to prevent war.

However, the challenge for history to contribute to world citizenship was more apparent only from 1950 onwards. This was evident in Education Pamphlet No. 52 which stated that its aim was to consider world history as an important dimension of history to teach world citizenship. It was indeed a movement which had already shown itself in its introduction into schools and public examinations. The history syllabi were no longer concerned mainly with British history, or with Commonwealth or European history, but with topics taken from world history, intended to encourage a better understanding of world affairs.

In fact, the main change that world history introduced was the attempt to make the world look like a whole. Topics of major civilizations in the past or of world significance were treated in their own right, free from bias and impartiality. This type of world history syllabi represented themselves, a new development in history teaching. This was the continuation of the link between history and citizenship education in the context of promoting international and world citizenship of the inter-war period. World history still remained in the history curriculum today. However, the emphasis was not on the teaching of world citizenship but on national history.

History as Social Education

Social studies movement in late 1940s and 1950s posed a challenge to traditional subjects for their contribution to citizenship education. Through rhetoric and pressure groups, it found itself a place in the school curriculum. Parallel to this movement, geography and history were also reflected as subjects responsible for citizenship education. For example, Norwood Report stated that

> "It is in such a treatment of history... that we believe the best contribution can be made in schools to the growth of an informed democracy... the instruction (i.e. citizenship) springs most naturally from the study of ordinary school subjects, particularly history...." (HMSO, 1943, p. 100)

Among historians, Burston argued that

> "there are few who would doubt that history can help to make our pupils better citizens. Some go farther, and say that if history is properly taught, no new subject, such as civics is needed." (Burston, 1948, p. 225)

Burston's conclusion was that

> "History imparts that touch of intuition which the sagacious citizen needs in handling the future, and we must never forget that the citizen's task is invariably to decide future policy, rather than to pronounce judgment on the past." (Burston, 1948, p.239)

While there was evidence supportive of history as a subject for citizenship education, history was also facing a tough challenge within itself and from social studies. In 1950, the Incorporated Association of Assistant Masters (IAAM) reported that the methods of history teaching had been attacked in the light of new ideas about the aims of education and about the content of curriculum. At the same time, it held firmly that history was

> "well fitted to exercise and improve certain skills of mind: weighing of evidence, detecting bias, appraising the probable, and separate it from the impossible, recognizing cause and effect; recommending it also for its capacity to enlarge sympathy and to develop a questioning attitude." (Betts, 1982, p. 11)

IAAM was, in a way, arguing for history's relevance for citizenship education. It stated that special attention should be paid to those branches of history which would promote interests in one's surroundings so that pupils would learn something of the problems of organized societies and of their political inheritance so that citizenship should not be merely negative and passive but alive. So long as the social purpose of history could be preserved and content and method of teaching for this purpose improved, there would be no danger for it to lose ground. But then history was open to the challenge of social studies --- an approach which was thought would better meet the needs of society. New attitudes toward education encouraged the expansion of material for history in terms of aspect (social and economic) and scale (world and local). History had already had too much subject matter to be covered in the school syllabi. Thus there arose the problem of selection and creation of space for the expanded matter - the materials for the transmission of citizenship values. Such problems of selection created opportunities to follow American examples whereby history was linked to social studies. In America, history had already been subsumed into integrated courses in elementary and secondary schools - mainly in social studies. But in Britain, the subject lobby was strong. History would not easily be dislodged from the school curriculum, but needed to respond to this challenge. Thus history began to change its content and aims to contribute to what social studies claimed as their area of citizenship education. When social studies began to lose ground in Britain in late 1950s, support for history to teach citizenship increased, leading to the New History Movement.

The New History and Its Link to Modern Citizenship Education

The New History argued for the inclusion of citizenship concerns in history lessons. By 1970, widespread change was evident in the teaching of history. The Historical Association sponsored the publication of Teaching History, a journal which could bring new teaching strategies to history teachers in all corners of Britain. The Schools Council set up two projects : History, Geography and Social Science 8-13 and History 13-16 in the early 1970s. Both of them initiated change in history teaching and provided supportive materials for the teaching of new history.

History, Geography and Social Science 8-13 project adopted a unique approach which favored the fostering of active citizenship. It looked at every teaching situation as characterized by the interaction between four variables - children, teachers, schools and environments. Of course citizenship formation was seen as the interaction of more than just these four variables. The following ideas taken from Spotlights connected it well to citizenship education:

1. the project emphasized critical thinking;
2. the project emphasized empathy;
3. the project's subject area was a sensitive area;
4. Place, Time and Society 8-13 must be closely related to the rest of the curriculum and to the rest of schooling;
5. the project emphasized interrelation rather than integration

The idea of interrelation rather than integration as reflected in this Schools Council History, Geography and Social Science 8-13 project provided the source of key concepts for the teachers to base their selection of content, as in the issue of empathy. It also demonstrated history and geography's relationships with other social sciences and how more effectively they could be used to foster education for modern citizenship. It also emphasized the fostering of critical thinking which was necessary for participation. It laid less emphasis on content and more on the process of learning. Thus teaching was through the inquiry approach and content was selected according to the educational objectives to be achieved and the historical skills to be acquired. In this way, history was made capable of enabling pupils to develop the various skills of active citizenship. It provided initiation for changing the link between history and citizenship from its more passive role to its more active role.

The History 13-16 project was another example of the new history movement which adopted a more progressive inquiry approach to tackle citizenship education. The project was really seen as to help teachers to help themselves after the publication of the article, "History in Danger" by Mary Price in 1968. This project aimed to encourage teachers to promote more pupil participation in their study of history and to foster modern citizenship - active and participating. This project published A New Look at History, which was well received by this movement. In addition, it came with four other titles meant for citizenship education. They were "Conflict in Ireland," "China," "Arab-Israeli Conflict," and "The Move to European Unity." Both political and controversial issues were discussed in these four books.

Developments in new history have also taken place in the universities where the content of undergraduate history courses began to change. The most significant swing was towards the modern period and towards world history. Blows found that the great majority of universities taught English and European history post-1939, and almost all offered courses on American, and nearly half on African, history. Asa Briggs pointed out that social history was the main area of development in history in the 1970s. Methods of teaching and studying history were also changing. These changes were translated to the school situation. During the 1970s, there was also a noticeable swing from preoccupation with the content of the history lesson to a concern for the ways in which children came to understand history. Thus the two Schools Council project discussed earlier both contained a chapter on the nature of history. Curriculum planning was objective-based, getting teachers to think carefully about what they were seeking to achieve and, in particular, it led to a heavier emphasis on deductive thought and skill acquisition, an important aspect of modern citizenship education.

The assumption underlying the earlier chronological outline syllabus, that the pupil could gain a sense of development and change, and a grasp of chronology was called into question, though not totally discredited. The History 13-16 Project, for example, adopted a 'study in development' as a part of its examination syllabus. Pupils studied the development of medicine from early times so that they may better understand 'the process by which change takes place in human affairs and continuities from the past survive into the present'.

As the chronological outline syllabus declined in popularity, so study in depth became more fashionable. The new history was reflected in the structure of history syllabus. The History 13-16 Project survey of 'O' level and CSE Mode I syllabi in 1971 and 1972 respectively found that, in both, 94% of pupils studied modern history either as British social and economic history, as British and/or European history, or as world history. Consequently, world history became popular in school in the 1960s. It continued to flourish in the 1970s, but faced competition for time from the various branches of social history, particularly local and environmental history, family history and historical demography.

There was also an increase in emphasis on using the activity method including simulation and drama techniques. Teachers favoring simulations had argued that they called on a range of skills, including the ability to use sources, communicate ideas, and appreciate other people's points of view, as well as providing the experience of making decisions. This strongly support the subject's claim of its capacity for citizenship education in this century.

Due to the fact that many of the ideas of the new history were not new, the subject continued to be fearful of losing its place in the school curriculum. Thus it still needed to commit more strongly to its duty in the transmission of citizenship values and skills. In relation to this, Elton explained that history teaching in school should not attempt to pre-empt what university courses offered, but should develop a kind of mind,

> "flexible and open to new ideas but at the same time capable of assessing them against the traditional, aware of mankind in its variety," (Elton, 1970, p. 226)

which would be valuable for the vast majority if not all school pupils whatever their future intentions might be. In this way, history offered a comfortable home for citizenship education. Because of this, Holloway and Heater both argued that an association between history and social studies was necessary. Heater suggested that the syllabus should be constructed either chronologically or on an era basis, and the material should be selected for its usefulness in exemplifying social science concepts, as well as for its inherent historical significance.

This association of history and social studies, had brought about a significant and lasting effect on the methods of the teaching of history in British schools. In the context of the development of and changes in citizenship education it was worth looking at since its main objective was to bring out the relevance of the subject for the practical needs of the pupils in their daily life and in their future life. The fact that the new history was more flexible than the old meant that it could create more opportunities for citizenship issues to be explored. Through the inquiry method, the children could acquire information and perceive relationships, draw conclusions and make decision. The mental training and the broadening of experience that were made possible was a fine education for young people. In this way, both the nature of history and the citizenship value of history were seen to be maintained.

After a decade of development, new history was officially questioned on the ground that 'real' history was being devalued through this approach. In many schools in the 1980s, history continued to be taught through a chronological or an era approach. Inquiry approaches did not take off widely. In many schools the emphasis remained on history for its own sake. Historical facts were stressed more than approaches. Its contribution to citizenship education thus tended to swing back to that of providing civic education. But it was at least generally different from the 'drum and trumpet' history of the earlier period. Some rhetoric and policies of history's preparation for its continuing contribution to citizenship education were still in place. For example, the document "History from 5-16" stated the aims of history teaching as

> "to contribute to personal and social education by developing certain attitudes and values; for example a respect for evidence; and toleration of a range of opinions;" and "to communicate clearly, employing a wide range of media." (HMSO, 1988, p. 3)

This document also made a number of claims of history's capacity to cover different areas of citizenship education and cross-curricular dimensions and themes. For example, it claimed that "history is well placed to enrich the school curriculum, and to prepare young people for life in contemporary society..." and it "has a particularly important role to play in preparing pupils to participate in multi-ethnic society...." Finally it emphasized that

> "a successful course in history ought to contribute towards the development of broadly educated people who are effective in their various roles as citizens, parents and contributors to the common good." (HMSO, 1988, p. 28)

In this way, it brought to light the continuity and change, and the rhetoric and policies concerning history's duty and capacity to teach citizenship education throughout the twentieth century.

Having said that, attention should also be drawn to Purkis's article in *Teaching History*, "The Unacceptable Face of History?" It uncovered some of the hidden agendas of history teaching, such as the continuing influence of R. J. Unstead, considered as the brand leader in school history. Though in the 1960s and 1970s, history had experienced rapid changes in favor of citizenship education, Purkis revealed that the most influential of the history school texts in primary and secondary education remained to be the old fashioned, Unstead's books, published in the 1950s. The following quotation supports this argument.

> "History teachers everywhere would agree that for more than twenty years our brand leader has been R. J. Unstead, some of whose early works the publishers A. & C. Black are now re-issuing. Although some of the material has been re-arranged, for example, making what was originally one book into two by adding photographs, maps, an index and sometimes a bibliography - the text remains substantially what it was in 1959... His approach is structured, safe and conventional, using a chronology that traditional teachers, especially those non-specialists teaching in primary schools, remember from their own schooldays. He emphasizes the long-running, happy and glorious success story of the great (white) British people." (Purkis, 1980, p. 34)

Today, citizenship education still has a place in history. Rhetoric and policies about it could be found, for example, in the National Curriculum documents entitled History Interim and Final Reports and the History Statutory Order of the 1990s. Though there was much early controversy, with the Secretary of State demanding more British History and giving a sharper focus for British experience, both Interim and Final Reports maintained a case for citizenship education. Four attainment targets were proposed to justify history's place in education, but the Statutory Order modified and reduced them to three. Citizenship education was there but not emphasized. Its implementation would therefore be left to chance. However, history non-statutory guidance document predictably and, in terms of official thinking appropriately, devoted some space to citizenship and cross-curricular themes such as:

> "National Curriculum (for history) relates to the main components in National Curriculum Council's Curriculum Guidance 8: Education for Citizenship, i.e. work, employment and leisure, the family and the nature of community." (HMSO, 1991, p. 15)

Apart from providing ample lip service to citizenship education, the statutory order did not, of course, restrict the potentiality of the use of history for teaching citizenship components. Thus, it remained positive that citizenship education could be taught through permeation method through history under the National Curriculum environment.

References

Blyth, W. A. L.(1973) Spotlights, Schools Council Project: History, Geography, and Social Science 8-13 (Univ. of Liverpool: School of Education).

Board of Education (1905) Suggestions for the Consideration of Teachers and Others Concerned in the Work of Public Elementary Schools (London, HMSO).

Board of Education (1923) The Teaching of History, Pamphlet No. 37, (London, HMSO).

Board of Education (1926) The Education of the Adolescent (London, HMSO).

Bourne, H. (1905) *The Teaching of History and Civics in the Elementary and the Secondary Schools* (London,Longmans).

Burston, W. H. (1948) The Contribution of History to Education in Citizenship, *History*, 33, pp. 225-40.

Cook, T. (1978) The Teaching of Local History in Schools since 1902, *History of Education Society Bulletin*, 22, pp. 42.

Dance, E. H. (1960) *History the Betrayer: A study in Bias* (London, Hutchinson).

DES (1967) Towards World History, Education Pamphlet No. 52, (London, HMSO).

Elton, G. R. (1970) "What sort of history should we teach ?" in Ballard, M. (ed.) *New Movements in the Study and Teaching of History*, (Temple Smith), 226.

Elliot, B. J. "An Early Failure of Curriculum Reform: History Teaching in England 1918-1940", *Journal of Education Administration and History*, 12, (2, 1980), 39.

Gould, F. J.(1927) *The League of Nations Spirit in the Schools* (London, Watts and Co.).

Hayward, F. H. (1910) *Day and Evening Schools: their management and organisation* (London, Ralph, Holland and Co.).

Hayward, F. H. (1942) *The Last Years of a Great Educationist: A Record of the work and thought of F. J. Gould from 1923-1938* (London, Duckworth).

Hughes, M. L. V. (1917) *Citizens To Be* (London, Constable).

Jones, R. B. (Ed) (1973) *Practical Approaches to the New History*

(London, Hutchinson).

Laurie, S. S. (1867) *History in the Primary School.*

League of Nations (1935) Second Session of the Advisory Committee on League of Nations Teaching, Bulletin of League of Nations Teaching, (2), p. 173.

League of Nations (1936) The Teaching of History and Geography, Bulletin of League of Nations Teaching, (3), pp. 16.

Minutes of the Wiltshire Education Committee (24 June, 1910), 16-17.

Pitt, G. (1893) *English History with its Wars Left Out* (London, T. Compton).

Purkis, S. (1980) The Unacceptable Face of History? *Teaching History*, 26.

Schools Council (1976) A New Look at History, History 13-16 project (London, Holmes McDougall).

Sir Cyril Norwood (1943) Curriculum and Examination in Secondary Schools (London, HMSO).

Strong, C. F. (1939) Training for Citizenship through History in: Education for Citizenship in Elementary Schools, (London, Association for Education in Citizenship).

Withers, H. L. (1904) *The Teaching of History and Other Papers* (Manchester Univ. Press).

About the Author

Jessie Wong Yuk Yong
Institute of Education
Singapore

wongyyj@nievax.nie.ac.sg

Jessie Wong is a lecturer at the National Institute of Education, Nanyang Technological University in Singapore. She received her M.Ed (1989) and PhD in Education (1991) from the University of Liverpool, UK. Her B.A. (Hons) Degree (1985) was obtained from the University of Malaya in Malaysia.

Dr. Jessie Wong has a broad experience in teaching, from elementary to tertiary level. Her current work as a teacher educator involves teaching, research, and curriculum design and evaluation. She has experience in international studies involving subjects from many countries in the world. Her current research interests are information technology in education, social studies and citizenship education.

Copyright 1997 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://olam.ed.asu.edu/epaa

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey@olam.ed.asu.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Marshall University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Rocky Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
Arizona State University

Dennis Sayers
University of California at Davis

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina--Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher Education

Mary P. McKeown
Arizona Board of Regents

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois--UC

Robert T. Stout
Arizona State University

archives | abstracts | editors | board | submit | comment | subscribe | search

EPAA
• enter the archives • browse the abstracts • the editors • the edit board
• submit article • submit commentary • search • subscribe
volume: • 1 • 2 • 3 • 4 • 5 • 6 • 7

This article has been retrieved 2218 times since July 4, 1997.

Education Policy Analysis Archives

Volume 5 Number 15 **July 4, 1997** **ISSN 1068-2341**

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass Glass@ASU.EDU.

College of Education

Arizona State University,Tempe AZ 85287-2411

Copyright 1997, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

Early Childhood Evaluation and Policy Analysis: A Communicative Framework for the Next Decade

Cynthia Wallat and Carolyn Piazza
Florida State University

Abstract

A major challenge for the next generation of students of human development is to help shape the paradigms by which we analyze and evaluate public policies for children and families. Advocates of building research and policy connections point to health care and stress experiences across home, school, and community as critical policy issues that expand the scope of contexts and outcomes studied. At a minimum, development researchers and practitioners will need to be well versed in available methods of inquiry; they will need to be "methodologically multilingual" when conducting evaluation and policy analysis, producing reports, and reporting their interpretations to consumer and policy audiences. This article suggests how traditional approaches to policy inquiry can be reconsidered in light of these research inquiry and communicative skills needed by all policy researchers. A fifteen year review of both policy and discourse processes research is presented to suggest ways to conduct policy studies within a communicative framework.

INTRODUCTION

Human development literature documents numerous attempts to stimulate debate, evaluation and recommendations concerning health and welfare policy and practice (Steele & Wallat 1997). Similar to other human service fields, it appears that policy dissemination is beginning to obtain a strong base in educational organizations and associations. The weekly newspaper Education Week has developed into "America's education's newspaper of record." The 20 thousand plus member American Educational Research Association (AERA) has ensured that dozens of educational policy presentations will be included in the annual meeting by advancing educational policy research to division status. Additionally, AERA editors recently devoted two issues of its premier journal, Educational Researcher, to include illustrations of building research and policy connections and the annual program was organized around the theme "Talking Together in Educational Research and Practice." Following this theme, the presidential address focuses on the questions, "How can we, as educators, create communication among professional educators, researchers and other public vested in education?" and "How can we explicitly convey a spirit of inclusion and discussion with all those interested in education? (AERA Annual Meeting 1996, p. 51). AERA Council members are also working on building a strong base by adopting a policy emphasis in the organization's strategic plan. These goals include devising mechanisms on communicating research in ways that influence policy and practice and engaging in specific activities to encourage research that is meaningful and relevant to practitioners and policymakers (AERA Council Minutes 1997, p.41).

Edward Lawlor (1996) suggests that "it is hard to imagine how this approach would not lead to a view of [policy] analyses as mere marketing" (p.114). The conception of policy analysis as "more communication" within and across "more communities" is a symptom of nadir of purpose and rationale and a warning sign of vulnerability, drift and low status (cf. p. 111- 112). It is not enough to say that policy analysts will gain power to make claims within policy arenas by publishing products that discuss research in simpler, more engaging terms. Such blanket statements do little to establish a connection to discipline premises regarding utilization of information or the dynamics of diffusion between sociocultural systems within contexts of tension (Wallat 1995). As Lawlor points out, to become a vital profession policy analysis must establish its connection to the intellectual traditions of the field and the "much larger movement in the social sciences to explore narrative and linguistic forms of social constructs" (p. 112).

The work presented here builds upon our initial identification of intellectual traditions in policy research on the relationship between language and policy (Wallat & Piazza 1991). We agree with Lawlor's analysis of what is needed to move beyond blanket statements about communication and dissemination. However, we disagree with his negative judgment that recent scholarship on discourses such as narratives represents a "shaky" contribution to policy research. To anticipate the conclusion, we agree with Lawlor's (and others) caution that if policy analysis is to become a vital profession it must create the capacity to lead its' members to perform their work in accord with the intellectual tradition of policy analysis as variable functions of communication, forms of persuasion , and modes of presentation (cf., Lasswell 1970; Wildavsky

1977). We disagree with his judgment that attempts to apply new knowledge in the disciplines of linguistics to analysis of achieving this endeavor are insubstantial. As Laswell (1970) has argued, the achievement of status and recognition of policy analysis as a profession in academic and public arenas depends upon critically examining its distinctive outlook. Such critical examination of policy analysis' distinctive problem orientation and social process (or contextuality) frame of reference requires continuous search for parallels between disciplines. As we have elaborated in other reports on types of discourse analysis, the disciplines of linguistics converge in emphasize on examining social contexts and social processes (e.g. Wallat & Piazza 1988, 1991).

The theme of language and policy relationships in this paper builds upon our earlier identification of intellectual traditions in policy research that have emphasized the need for policy scientists to pay attention to variable forms of communication and community in the production of final reports. We illustrate here the relationship between language and policy by considering communication functions and tasks undertaken before the production of final policy reports. Given our professional interests in early childhood, we eventually narrowed a fifteen year set of examples of communication functions in policy analysis to early childhood policy studies. Our approach essentially follows Marshall and Rossman's (1989) advice to identify and name a process and construct and explore its attributes in a variety of functions. First, we decided to concentrate the review on illustrating the process of communicative functions as exemplified in AERA's policy journal called Educational Evaluation and Policy Analysis (EEPA). Second, we identified a set of language in public life constructs from multiple discipline studies of forms of communication and community. The illustrations of variable communication functions in policy analysis traditions then serve to point to connections to knowledge on linguistic forms of social constructs. The fifteen year review may also serve to suggest connections that need to be made to expand the scope of early childhood contexts and outcomes to address critical human development policy issues.

SKILLS NEEDED BY POLICY RESEARCHERS

It has become commonplace to expect that students of human development need to be comfortable with the basic concepts and underlying assumptions of social sciences inquiry in order to judge the results of early intervention. However, developmental researchers have argued that professionals are increasingly obligated - by scientific considerations, by societal demographics, by professional societies, and by moral persuasion - to address how their fields of developmental study can draw more direct links between their work and policy questions (MacPhee, Kruentzer & Fritz 1994, p. 713).

Child and family policy experts also point out that expanded skills are necessary in order to help others decide which promises are reasonable for inclusion in public policy initiatives (cf. Datta 1983, p. 469). Datta (1982) predicts that just as methods used to disprove hypotheses drew "the best and the brightest" in the past, the use of approaches to explain what is tenable in development programs and policies research can attract a new generation of researchers who are "methodologically multilingual" (p. 144). As Lazar (1983) reminds us, the ground has been broken in evaluating the enduring effects and the practical economic implications of intervening in the contexts of children and families. At issue is how human development fields will face their diminishing ability to use classical analysis technologies in the light of social change and community sophistication. The scope of concern is reflected in expansion of development and policy agendas such as, (a) How and why some children and families fare better than others? (b) How and why a social structure, position, or system could or should affect the individual ? (c) Who benefits from early childhood programs? (d) Are some interventions more successful with some individual and social group capabilities acquired in social life? and (e) Where should scarce resources be allocated? (cf. Huston, McLloyd & Coll 1994; Hymes 1974; 1979). The technical tasks of developmental researchers will be to outline what advances are possible to available evaluation and policy analysis approaches in order to satisfy the expectation that investigations can draw more direct links with policy questions (Huston, McLloyd & Coll 1994, p. 282). Their theoretical task will be to become familiar with evaluation and policy analysis topics and approaches that have been undertaken within anthropology, economic, social psychology and sociology perspectives (Susman-Stillman, Brown, Adam, Blair, Gaines, Gordon, White & Wynn 1996, p. 7).

PERSPECTIVE

It is no small task to become methodologically multilingual in order to consider critical social issues in a broader framework. Datta (1982) reminds us that the conduct of, and completion of, multimethod studies involves a considerable range of purposes, sequences, and functions. Distinctions of potential significance which have helped awake the nation's conscience to situations of displacement, life on the margins, and blocked possibilities (Note 1), are dependent upon communicating within both academic and political arenas. For example, Popkewitz (1991) points out that it is shortsighted to believe that the right mix of availability of public funding for intervention research -- and a group of researchers practiced in multiple approaches driven by competing assumptions about classical models of the change process--will be able to have an enduring effect on social, health and educational policy.

Popkewitz reached this conclusion after examining reform efforts in the United States since the 1960s through an historical, philosophical and social theory framework. However, he also believes that drawing links between development research and policy questions might be able to have three long lasting contributions : (a) dissemination of the knowledge of disciplines along with the basis of these endeavors (i.e., making visible variable approaches of arguing, thinking, and "seeing" the world), (b) recognition that discipline knowledge produces a curious anomaly of explicitness and ambiguity (i.e., making visible the production of discipline knowledge as a "process that naturally involves ambiguity, tentativeness, and inventiveness as core dispositions" p. 140), and (c) advances in understanding of the functions of multiple inquiry approaches (i.e., making visible the link between policy research and policy proposals as social practices and social regulations of inquiry).

Popkewitz conceptualizes these three examples of long lasting contributions of development and policy research to demonstrate his concern with the need to expand support for development of social epistemology as a legitimate art of inquiry. As explained in his proposed options for considering the relation of researchers and social movements, social epistemology would consider the objects constituted as "the knowledge of policy" as a central focus. Within this focus, "the knowledge of policy" becomes a social practice amenable to inquiry; statements and words included in reports "are not just signs or signifiers that refer to fixed things, but are forms of social practice" (p.219). This conceptualization of the relationship between knowledge (epistemology) and power is intended as a means of making research practices such as evaluation and policy analysis accessible across multiple disciplines. As Wildavsky (1977) argues, "Analysis should be shown not defined" (p.10). Only through juxtaposition of available and manipulable problem solving means and resources can analysis be shown. "Who are we to say which field - law, sociology, social psychology, philisophy, history, anthropology, or whatever - might make the best contribution to policy analysis?" (p.12).

The next sections provide a starting point for venturing into the study of social practices of policy research. Overall, we address the possibility of alternative conceptions of policy research through presentation of a communicative framework for considering multiple approaches to evaluation and policy analysis. The nature and substance of approaches, considered as forms and functions of communication, include accountability, case study, discrepancy, ethnographic, experimental, expert opinion, illuminative, judicial, naturalistic, and responsive. The framework is based upon the assertion that what is--or will be--tenable is a constant in policy studies (Datta 1983). The idea that better knowledge produces better decisions is misleading. The types of skills and training needed to address critical policy issues require considering how a combination of policy approaches, or functions, may eventually lead to modest policy claims, contingent policy claims, or strong claims that warrant funds for continued investigation.

THE CHALLENGE

The question of what is tenable involves a number of issues. (Note 2) Current accountability rhetoric suggests that providing convincing evidence of the "enduring effects" of public policy initiatives must include development of conceptualizations of educational and other social development programs as a means of initiating a reinvestment spiral from limited initial capital. (Note 3) The widespread adoption and use of metaphors such as "enduring effects," "opportunities for success," and "practical economics" by legislators means that development professionals must communicate positive accountability indicators, indicators that go beyond the repertoire of terms heavily relied upon in the discourse of psychometric research. Evaluations must demonstrate how programs both provide for local resources management and human resources management, employing relevant strategies of cultural, political or religious revitalization (Easton 1989). To uncover and explain such post early development effects requires that development professionals be well versed in traditional and evolving approaches to evaluation and policy analysis. Easton (1989) refers to this job skill as paying increased attention to "the notion of integrating the educational field," that is, paying increased attention to the multiple approaches to inquiry that have developed in the social sciences to help relate education to the rest of the social system (p.438).

VIEWING RESEARCH APPROACHES AS DISCOURSE FUNCTIONS

Addressing the growing number of calls for "multimethod" evaluation and policy analyses (e.g., Green, Caracelli, & Graham 1989; Caracelli & Green 1993), does not have to start from scratch. Recent analysis of a decade of contributions to the communicative functions of policy analysis has highlighted convergence with several decades of advances in understanding the semantic and pragmatic aspects of communication and the multiple communicative resources individuals have at their disposal (Wallat 1984; 1991a). For example, the development of detailed procedures in descriptive linguistics since the 1960s has resulted in empirical explorations of multiple communicative resources which encompass notions of negotiation, accommodations, and pluralism.

Lakoff's (1990) formulations of language use as power and language resources as a composite of the three properties of form, function, and meaning, is especially valuable in two ways to policy analysis. First, such formulations about language resources provides a theoretical framework for acting on 25 years of advice by policy researchers to pay attention to language (e.g., Benveniste 1991; Smith 1992; Wallat & Piazza 1991). Second, advances in understanding the semantic and pragmatic aspects of oral and written communication are especially valuable in meeting the responsibilities of evaluators. At a minimum these responsibilities include awareness of particular applications of language in accomplishing everyday events, in accomplishing variable functions of the social sciences as academically institutionalized inquiry, and accomplishing variable functions of policy directed work in the contexts the researcher hopes to study (cf. Borich 1983). As Lakoff points out, "the trick for all of us is to grasp the generalizations, the larger picture. What is the connection between the form [function and meaning] of a communication and the power it provides its' user.... All language is political; and we all are, or had better become, poiiticians" (pp. ix, 2).

SCOPE OF THE REVIEW

The sections that follow deal largely with assertions about language resources by demonstrating the contributions of discourse process research to policy work. Section one sets the stage by presenting a 15 year compilation of policy analysis approaches in terms of their primary communicative functions. As elaborated upon in section two, the concept of functions of language is one of 9 key linguistic concepts that parallels the theoretical and practical interests of policy work. One purpose of the policy work compilation in Section one is to establish a connection to the intellectual traditions of the field and the movement in the social sciences to explore linguistic forms of social constructs. Demonstrating parallels between social practices and inquiry functions across disciplines is an untapped resource for individuals as they move towards the complex role of a full- scale policy scientist who is knowledgeable of the policy process and knowledgeable in the policy process (Lasswell 1970). An example of one scholar's move towards policy research points to the benefits found in becoming aware of types of discourse analysis available. Conquergood's (1991) work in international war zones and inner city Chicago led him to explore historically the disagreements that are part of the organization of social sciences' struggle with "deep epistemological, methodological, and ethical self-awareness" (p. 179). His identification and analysis of such parallels across the social sciences led to becoming reflective about the kinds of knowledge and their attendant discursive styles that get reconstituted and privileged in social science research propositions and research practices. "And, most importantly for critical theorists, what configuration of socio-political interest does a disciplines' scholarship serve" (p. 193).

Section one illustrates some first steps that can be taken to address some of the scrutiny issues that need further reflection in the latest wave of attempts to recalibrate research with a focus on social practices (cf., Cohen & Garet 1975; Rein 1983; Schubert 1980). The section includes an introduction to ten classical evaluation and policy analysis approach categories; it outlines discourse concepts for building a communicative framework for considering these categories in the remainder of the article. This section one outline serves as a strategy to organize references to fifteen years of educational evaluation and policy analysis studies that used these approaches and present concepts from research on discourse processes as a new way to think about evaluation and policy analysis. The policy research reports that are included as illustrations of the communicative functions of policy analysis were selected based upon two criteria. First, the authors of the reports included statements which referenced how they intended to build upon one of ten most frequently reported approaches in the American Educational Research Association's (AERA) policy journal which has recently completed its first 15 years of publication (i.e., Educational Evaluation and Policy Analysis). The second criteria for review and selection of 46 early childhood policy reports was that the authors included elaboration of the functions and procedures of the approach they used.

The discussion of projects included in the fifteen year compilation highlights a point of convergence across the concerns of evaluators and policy analysts, development and discourse processes researchers, and critical theorists: the need to reconsider how social sciences research approaches are capable of serving the social-communicative function of expanding the classical conceptual and methodological focus on institutions and organizational goals to include individuals and their beliefs, interactions, interpretations, and alternative reactions and responses. Having related early childhood research with intellectual traditions in educational policy work as well as sociolinguistic work, section two suggests ways of enhancing this relationship. It reintroduces the nine communicative framework concepts that may help child, family and community researchers include attention to discourse processes as they keep abreast of new policy work and/or participate in evaluation and policy projects. From a communicative perspective, the framework supplies a central motif for considering how the next generation of analysts from multiple fields of developmental study can draw more direct links to policy issues by: (a) learning to define policy work as a social event that serves communicative functions, and (b) learning to become methodologically multilingual through the use of metacommunication tools throughout the research cycle.

The third section illustrates how a communicative framework may be maintained in future evaluation and policy work by presenting a sample of procedural details from the research approach reports compiled in section one. The procedural details are offered as examples of the nine language and policy concepts--or metacommunication tools-- introduced in section two. Following sociolinguistic analysis aims, the purpose of using these concepts to scrutinize and design policy studies is to keep ourselves and others appreciative that "language is not just something that shapes our understanding of the world, but also a tool by which we can discuss and evaluate these same understandings" (Nielsen 1995, p.11). We introduce and explicate the concepts in this article as a strategy to help the next cadre of development and policy researchers develop an initial assessment of the conditions of possible communication in the design and conduct their policy studies. Once such conditions are identified, it will be possible to address the larger issue of what we do with our research material and how we represent human action and context (cf., Grimshaw 1987; Popkewitz 1991; Risemann 1993; Wallat 1991a; 1991b).

LINKING EARLY CHILDHOOD RESEARCH, EDUCATIONAL EVALUATION AND POLICY ANALYSIS, AND LANGUAGE WORK

Early childhood research has a long history in the United States. Recent reviews acknowledge over 60 years of efforts by organizations such as the Society for Research in Child Development to contribute to the development of policies that influence the lives of children and their families (e.g., Stevenson & Siegel 1984; Wallat and Steele, 1997). One indication of how educational researchers have included consideration of early childhood evaluation and policy analysis can be found in the content of reports published in the first fifteen years of the AERA journal which was established to professionalize the subject of evaluation and policy analysis.

Tables 1 and 2 demonstrate that locating multiple functions of early childhood evaluation and policy analysis approaches is not difficult. The educational evaluation and policy analysis approach categories listed in Table 1 represent a compilation of over 130 policy reports which have been published since 1979 in EEPA. Matching the sample of articles to the labels used to flag ten approaches used by contributors to EEPA was quite straightforward: the approach type was often included in the title of the article, and / or the researchers included statements in their EEPA articles to introduce these generally known labels of the functions served by these approaches (cf. Stake 1981).

Authors of the evaluation and policy studies used to compile the numerical chart in Table 1 are identified in Table 2 along with the topic studied. Again, these references are organized according to the particular approach category the authors used to flag their project findings or approach discussion. Table 2 also identifies the authors of 46 evaluation and policy articles concerned with early childhood. To minimize the arbitrary nature of most categories used in compilations or reviews of reports, we relied on the emphasis each author presented in their article to decide about inclusion of reports into the approach categories labeled accountability, case study, discrepancy, ethnographic, experimental, illuminative, judicial, naturalistic, and responsive. If the author(s) gave attention to more than one approach, the article is cited more than once.

TABLE 1

Functional Targets of Analysis of Ten Approaches Used in Educational Evaluation and Policy Analysis (cf, Stake, 1981b)

1.1. Collection of policy reports highlighting the following 10 discourse functions:

accountability	to assure laws, policies, rules, contracts, are honored
case study	to concentrate on a single case and its complexity
discrepancy	to emphasize formal objectives and their contrasts and impact
ethnographic	to emphasize cultural relationships
experimental	to report "controlled" treatments that aim toward assessment
expert opinion	to organize information around providing feedback for decisionmaking
illuminative	to portray situations as readers themselves may see things
judicial	to optimize presentation of the cases for and against
naturalistic	to observe ordinary events in natural settings
responsive	to fix on meanings and concerns held by key constituencies

1.2 Total Number of Approach Discussions/Examples

Approach	Number of Discussions of Approaches	Number of Early Childhood Discussions Reviewed (1979-1994)
accountability:	17	10
case study:	17	12
discrepancy:	5	3
ethnographic:	8	1
experimental:	11	9
expert opinion:	2	2
illuminative:	5	1
judicial:	9	5
naturalistic:	8	2
responsive:	5	1
TOTAL	87	46

TABLE 2

2.1 Authors Cited in the Review and Topics Covered In Articles

Approach	Authors	Topic
accountability:		
	Barnes & Ginsburg, 1979	Title 1
	Beneveniste, 1985	Design of Statewide Accountability Systems
	Cohen, 1979	Non-public institutions
	Cross, 1979	Title 1
	David, 1981	Title 1
	Dougharty, 1979	Expenditures per student
	Felter, 1986	CA Public Schools
	Gorwin & Green, 1980	Linguistic forms in which value terms appear
	House, 1979b	Follow Through
	Kean & Scanlan, 1979	Title 1

	Nations, 1982	State-funded Compensatory Systems
	Slavin & Madden, 1991	Chapter 1
	Smith, 1982	State Departments of Education
	Stonehill & Groves,1983	Title 1
	Stufflebeam & Webster,1980	Conceptualizations of Accountability
	Vander Ploeg, 1982	Title 1
	Warfield, 1994	Chapter 1
Case Studies:		
	Alkin & Daillak, 1979	Utilization of case studies
	Barnette, 1983	Interagency Linkages
	Bissell, 1979	CA Early Childhood
	House, 1979b	Follow Through
	Kirst & Jung, 1980	Title 1: 13 years
	Mazzoni & Clugston,1987	MN K-12 Reform
	Mercurio, 1979	Co-operative K-6, decison-making
	Monti, 1979	Desegration policies
	Porter, 1980	Studies of Educational Innovations
	Sadler, 1985	Site Selection Guidelines
	St. Pierre, 1982	Follow Through
	Ball, 1990	2nd Grade Mathematics
	C de Baca, Rinaldi, Billig & Kinneson, 1990	Chapter 1
	Cohen, 1990	2nd Grade Mathematics

	Cousins & Earl, 1992	Elementary School
	Peterson, 1990	2nd grade mathematics
	Winfield, 1991	Chapter 1
Discrepancy:		
	Birman, 1981	Title 1 and Pl 94-142
	House, 1979a	Report Narratives
	Le Mahieu, 1984	2nd grade mathematics: test instruments
	Osterlind, 1979	Utilization of evaluations
	Wehmeyer, 1979	Program participants' Views of an Early Child Education Program
Ethnographic:		
	Dorr-Bremme, 1985	Theory in evaluation fieldwork
	House, 1979a	Report Narratives
	Koppelman, 1979	Teacher-centered evaluations
	La Belle & Moll, 1979	Preschool Bilingual
	La Compte & Goetz, 1982	Guide to doing ethnographic evaluations
	Noblitt, 1984	Contributions of applied ethnography
	Stake, 1979	Use of subjective methods
	Wirt, Mitchel & Marshall, 1988	Analyzing values in state policy systems
Experimental:		
	Barnett, Freed, Mobasher,1988	Presch. Curr. Efforts
	Cooley & Leinhardt, 1980	Grade 1&3 Reading & Mathematics

	Darlington, 1981	Pooling data on pre- school
	Eisner, 1979	Qualitative and Quantitative
	Ellwein, Walsch, Eads & Miller, 1991	Kindergarten
	Gersten, Carnine, Williams, 1982	K & 1st Grade Teaching
	Health & Brandon, 1982	Special Education school programs
	Linn, 1979	Title 1
	Odom & Ferell, 1983	Modifications in 20 ECE designs
	St. Pierre, 1979	Follow Through
	Stufflebean & Webster,1980	Conceptualizations of experimental approach
Expert Opinion:		
	Klein, Gold, Stalford, 1988	"Convening process" re Title 1
	House, 1979b	Follow Through
Illuminative:		
	Crain & Mahard, 1979	School Desegration
	Escobar, Barnett & Keith, 1988	Parents willingnes to pay for Early Childhood Special (ECSE) Services
	Florio, 1979	Utilizations of Evaluations
	Murphy, 1980	The state role in education
	Shapiro, Secor,& Butchart, 1983	Management Training Programs
Judicial:		
	Jordan, 1981	Applications of the Judicial Process
	Kirst & Jung, 1980	ESEA: A 13 year view
	Meisels, 1985	ECSE State of MA Reg.

	Nations, 1982	State funded Compensatory Programs
	Swan, 1984	ESCE Across-state Regs
	Thompson, 1981	Utilization of research in utilization
	Turnbull, 1982	Legal Antecedents to Mainstreaming
	Wolf, 1979	Pl 94-142 Regulations
	Wood et al, 1986	Model for identifying gifted program needs
Naturalistic:		
	Barnette, 1983	Interagency Linkages
	Cooley & Leinhart, 1980	Grades 1&3,Rdg.& Math
	Door-Brenne, 1985	Theory in Evaluation fieldwork
	Eisner, 1979	Qualitative & Quantative
	LaBelle, Moll & Weisner, 1979	Preschool Bilingual
	Lynch, 1983	Qualitative & Quantitative
	Smith, Dwyer & Prunty, 1981	Contributions to Problem-solving knowledge
	Williams, 1986	Evaluation Standards & Naturalistic Criteria
Responsive:		
	House, 1979a	Report Narrative
	Maxwell, 1984	Rating scale for different types of reports
	Schermerhorn & Williams, 1979	Utilization of different types of reports
	Stake, 1981a	Evaluations as persuasions
	Stake, 1979	Use of subjective methods
	Wolf, 1979	PL 94-142

DISCOURSE CONCEPTS FOR ENHANCING POLICY WORK

This article's overview of educational evaluation and policy studies is intended as an introduction to the notion of using discourse processes and policy concepts as a metacommunication tool by which development researchers can discuss and evaluate how the 10 approaches and scope of work presented in Table 1 and Table 2 can be thought about in new ways; ways that highlight new ways to think about what we do with our research materials and what we do as we work with informants.

We believe that the building blocks of such advances are available. Educational policy researchers have suggested for some time that evaluations should be taught as instances of information strategies that may help "bring elements of documented fact and empiricism to the political process" (Bissell 1979, p. 37). At issue is demonstrating how classical approaches can be enacted to take advantage of --- and to portray -- the multiple social discourses and power relations which constitute the nature of society and its problems (cf. Cohen & Garet 1975, p. 17). The concepts presented in Table 3 were chosen in order to encourage readers to use a communicative framework to evaluate how evaluation and policy analysis approaches can meet the emergent demands for mixed - method designs and a participatory policy discourse that considers insider's account of events, objects and action.

Table 3

Contextual Concepts Used in Studies of Forms of Communication and Community and Recommended for Considering Language and Policy

Functions of language	Sociolinguists have proposed that the oral and written structures of an individual's language should be viewed as representing social behavior; as "meaning potential"; as functional options chosen among the multiple linguistic resources all writers and speakers have; and as realization of the social context at a particular time and place. Such a linguistics, or study of language-in- public- life, will have an essential property. Its practice and theory will be adequate to all the means employed in oral and written language and all the meanings inherent in choice of dialect, variety, style, conversation or narrative genre, or the occasion to speak or be silent. (Halliday, 1973)
Language per se is ambiguous	The term ambiguous is used in the sense of evolving definitions of the meaning of social situations, or participant structures, or enactment of routines. Ambiguity is especially visible in situations in which rules and expectations are not verbally stated or remain implicit in descriptions of writing. (Wallat & Green, 1982; Wallat & Piazza, 1991)

Participant Structures	It is usually possible to use language to recreate aspects of social structure: to identify the particular way individuals arrange communication and community, or, in other words, to determine how involvement has been allocated, ended, or recommenced, or topic shifted. (Philips, 1993; Cicourel, 1994)
Communicative competence	The concept of communicative competence refers to knowledge of cognitive and social conventions that speakers and writers use as a resource to initiate and sustain involvement. In studies of oral language-in- public-life, the process of establishing a frame refers to the efforts individuals spend on reaching agreement on what activity is being enacted and how it is conducted. (Gumperz, 1982a)
Interpretative frames	Although we lack the conceptual tools for removing all sources of ambiguity in communication, we can begin to learn to ask ourselves two questions: How adequate are the grounds for our interpretation of the phenomena of behavior under study? and, more specifically, Have we adequately considered the multiple meanings that a surface language form can have to the listeners, or readers, of a report? (Cazden, 1986)
Schema/Frames	Although the concepts of schema and frame have a range of definitions, the definition of schema as "knowledge structures reflected in speech and writing forms of participants --sets of expectations that people have for language, settings, and the structure of interaction" is the definition suggested here. The definition of frames that is also useful for considering information processing as a cognitive and social process, is frames as the superordinate message (or metamessage) about what is actually being engaged in, and what changes of frame may be appropriate. (Tannen & Wallat, 1983)
Construction of Social Norms	The concept of construction of norms as group process has been proposed to account for how individuals negotiate a complex system of ambiguous rules in social settings, and/or participant structures. The form of communication and community can be linked to meanings participants evolve based upon the wider social order, social setting, routine, and interaction frame. (Handelman, 1975; Wallat & Green, 1982)
Enactment of Routines	The concept of enactment of routines was originally introduced to suggest how organizational behavior can be understood less as a set of "acts," "choices," and searches for "reasons" and "motives" and more as consequences of an organization's function. (Mehan, 1984)

Style options/shifts	Turns taken in participant structures; norms constructed, modified, or suspended during participant structures; and routines enacted in institutional settings, are all made up of basic tools with which people communicate. Style refers to each person's decision about which communicative devices to apply, and to what extent in a given situation. (Tannen, 1984)

As demonstrated in Table 3, much has been learned in the last two decades about oral and written discourse from multiple disciplines (anthropology, cognitive psychology, linguistics, social psychology, and sociology.) The sample of new constructs provide a communicative framework to meet the recognized need to build upon research utilization principles and help identify and describe those hard things to document: (a) how findings from studies of early childhood social programs "percolate" into people's awareness; shift perceptions of what counts as effects; affect the climate of opinion; and develop a "mosaic of understandings" about early childhood educational practices (Bissell 1979, House 1979, Weiss 1980), (b) how policy studies fit into an individual's interpretative framework (Sadler 1980), (c) how policy studies may affect a series of small and uncoordinated actions, foreclose some options, open perusal of others, and, eventually, result in hard to track, diffuse and indirect social problem-solving and decision-making events (Weiss 1980), (d) how policy studies can bring to the forefront of social science research clearer understandings of aspects or features of circumstances, components, conditions, dimensions, episodes, events, factors, items, qualities, units and the like which researchers use to capture the notion of social context (Wallat 1991b).

By focusing on language and policy concepts to review the first fifteen years of EEPA work, the sections which follow attempt to provide a language and policy resource (i.e., "metacommunication concepts") for helping readers become methodologically multilingual by becoming interested both in the strategies of particular [discourses] and in those shared by all discourses (Lakoff 1990, p. 5). The concepts outlined in Table 3 are a step in this direction. Working from behavioral records of participants' interactions across institutional settings and events, discourse researchers have, by explicating these nine concepts, put developmental professionals in a good position to begin to "see" more clearly the nature, structure, and function of the discourse demands that are part of the day-to-day processes and products accomplished during an evaluation and policy study.

SECTION ONE SUMMARY

The fact that the discussion of early childhood research and policy work published in the first fifteen years of EEPA can be cast in discourse research terms stems from the notion of educational evaluation and policy analysis approaches as discourse functions (Wallat & Piazza 1991). When discussed within a communicative framework these approaches can be reconsidered as social inventions which have been created, modified, expanded, or transformed to meet variable and shifting political and scholarly interpretations of what educational evaluation and policy analysis should be. Treated as social inventions that attempt to accomplish discourse functions, evaluation and policy analysis approaches can be considered as resources in addressing the call for a multimethod conceptual framework in development studies. Discourse conceptualizations, such as Lakoff's model of language resources as a composite of the properties of form, function, and meaning serve to help researchers see these links in their day to day project tasks. Given that language researchers conceptualize discourse functions as the use of a variety of cognitive and social processes, and the selection of oral and written dimensions which can be precise and indefinite, referential and expressive, and quantitative and qualitative, the next generation of policy inquiry can consider variations in public discourses and leave behind unresolvable debates about why academicians can not communicate with policymakers. If policy analysts take the time to become familiar with the features of academic, bureaucratic, and legal language the field of educational research may eventually move away from blanket statements such as more clarity and brevity will lead to more communication.

Lack of utilization of educational research is only incidentally a miscommunication problem. Variations in discourse features can be politicized into borders which define different people as antagonists. In other words variation in discourse features across social settings can be viewed as a resource for mutual exploration of social issues as well as a resource for conflict (McDermott & Gospodinoff 1976).

USING A DISCOURSE FRAMEWORK FOR EVALUATION AND POLICY ANALYSIS

Discourse concepts can be used by evaluators and policy analysts to enhance their profession's accomplishment of three tasks: (a) considering what gets defined and who does the defining, (b) contributing to qualitative insights, and (c) articulating variable interpretations of policy contexts and outcomes (cf. Lakoff & Johnson 1980; Gumperz 1982; Cronbach 1975). Policy researchers may find that the concepts are useful in day to day work on a variety of projects. For example, as individuals involved in a particular project begin to work with each other, they begin constituting events in the formulation, design, and conduct of a policy study. While some may consider such events as simply routine, discourse research provides insight into the resources individuals rely upon to define the situation and establish their work within participant structures, as well as insights into the social and cognitive processes embedded in the enactment of routines.

The nine concepts previously defined in Table 3 are used in discourse studies to identify and explain variation in functions, forms, and features of communication across home, community, business, school and other institutional settings. On a cautionary note, evaluators and policy analysts unfamiliar with sociolinguistics research may find that prior notions about language, social organization, or development interfere with their ability to understand these nine concepts. While public media and workshop brochures offer exhortations on a perceived decline of clear writing and speaking, discourse researchers provide representations of oral and written language in quite different terms. To capture the indefinite and precise features and resources of language, discourse researchers use descriptors such as inherent ambiguity (e.g., Crystal 1980); inherent indeterminacy (e.g., Dore and McDermott 1982); inherent variability (e.g., Hymes 1979); and inherent, inextricable links between the referential and social aspects of meaning in language (Erickson 1978). With these notions in mind, discourse researchers would tell an audience of policy researchers that even a hypothetical society of all standard language speakers and writers would show detectable variations in expectations about language fluency, expressiveness, and use of styles (cf. Becker 1988; Hymes 1979).

Individuals and social groups have constructed a variety of conventions to deal with the resource and constraint features of language. Such conventions are visible in individuals' and social groups' ideology and values regarding communication standards; what is considered "clear and precise" in one code may not be considered "clear and precise" in another's code of socialized correct conduct. Therefore, a key to understanding the organization of -- and social practices of -- oral and written discourse functions in a particular policy project is to learn how to become aware of the "rules" or conventions clients and audience have about the use of language. Hymes (1979) refers to this as undertaking observation of an organization of these means, roughly a "what" and a "how."

> A key to the organization of language in a particular culture or period is restriction of free combination of "whats" and "hows," the things that must be said in certain ways, the ways that can be used only for certain things. The admissible relations comprise the admissible styles. In effect, the study of language is fundamentally a study of styles. (Hymes 1979, p. 8)

In other words, in order to begin to "see" the inherent characteristics of oral and written texts in operation, to "see" how situations, roles, and activities affect and shape ways of using language, we need to go beyond questions of rules of grammar, or rules of production, in the design of educational evaluation and policy analysis. Lasswell (1970) anticipated this problem and advised policy researchers that their well established categories and neologisms in formulations of policy and society did not have to be dropped. "In the future [policy researchers] can re-edit [their] terms in ways that increase the interconvertability of jargon systems" (p.16). More recently the work of students of discourse systems have added further specifics to Lasswell's premises on thinking and talking about policy and society. What is required is awareness of the social meanings attached by individuals or groups to variations in oral and written styles and an awareness of individuals' or group attitudes or notions about language. The definitions in Table 3 are a beginning.

Functions of language. The question of how individuals communicate information and persuade others in actual situations is still far from being resolved. Some have observed that "it is not words which mean things but [individuals] who, by words, mean things; that a statement does not represent a fact but that [individuals], by a statement, mean facts" (DeMauro 1967, p. 2). Yet discourse researchers argue that finding answers to questions about how information and persuasion are used to create certain rhetorical effects and how they can be analyzed requires paying attention to how ordinary forms of communication are empty by themselves. "As in architecture, form is function, and is meaning as well" (Lakoff 1990, p. 27). They attempt to persuade human service professionals that it is possible to develop understanding of how day to day talk and writing is a composite of all these aspects. Lakoff (1990) recommends that we begin to think about what we read and hear from the vantage point of a schema/frame she refers to as a three-sided triangle. She uses the concept of schema/frame in terms of its usefulness as a metacommunication tool; the metaphor "three sided triangle" can guide our observations about how functions of language are a reflection of beliefs about speaking and writing in the policy contexts we are studying. For example, if we are conducting an "accountability" study, we need to become aware of the meanings that the contract monitors as well as intended informants and report audiences associate with the abstract concept "accountability." Lakoff's suggestion of focusing on a three-sided triangle is a reminder to ask ourselves , How are the forms of language related to the uses we and others make of them? (p. 6). Hymes' (1972) contends that such questions provide a primer for observing pluralism in day to day actions. People who know the same sounds, words, and syntax may not know [other individuals'] rules for interpreting [such sounds, words, and syntax] as requests or commands; [as signals] for the topics that can be introduced ...; [as signals] for taking turns and getting the floor; for making allusions; avoiding insults, showing respect and self-respect in choice of words, etc. (Hymes 1972, p. xxxviii). Mutual understanding depends as much upon common linguistic means, in the narrow sense, but also on being aware that reaching understanding involves negotiation about how those who want "more communication" reach agreement about ways of using and interpreting speech.

The point both Lakoff and Hymes make about linguistic means is that speakers and writers already have resources they can use to begin to analyze these facets of public discourses more precisely. Policy research critics such as Garnet and Cohen (1975), House (1979a), and Sadler (1985), argue that educational policy researchers might pay attention to the repertoire of choices speaking, listening, writing, and reading communication resources comprise. To date discourse researchers have illustrated how variation in form can begin to be thought about through considering a relatively few general categories of functions of language (i.e., getting things done; controlling the behavior of others; telling about oneself; using forms of personal expression and social interaction; finding out about things; communicating something for the information of others; and, relating to contexts of use as speaker, listener, writer, and reader) (cf. Wallat 1984, p. 24-25). Essentially, those who speak and write for public audiences are attempting to solve the problem of creating what discourse researchers refer to as the 'textual' function, whereby [oral and written] language becomes text, is related to itself and to its context of use. Without the textual component of meaning, we should be unable to make any use of language at all. (Halliday 1973, p. 44)

Halliday's (1973) definition of functions of language in Table 3 is essentially a proposal to adopt a functional viewpoint and, thereby gain ideas about how "meaning is related both to the internal structure of language and to the social contexts in which language operates" (p.8). Functionally the choice of a word or phrase "may have one meaning, its repetition another and its location in structure yet another" (p. 109). More specifically, there is a professional development benefit which is derived from awareness of the functions and formats of special languages or codes across participant structures; it is an appreciation of the multiple resources which results from the variations possible in both oral and written discourse forms, and the resources we can use to meet the myriad demands made on participants' communicative competence (cf. Cazden 1986, p. 437).

Language per se is ambiguous. Discourse researchers have presented some interesting examples of miscommunication that can be traced to the ambiguity of language. Of particular interest to readers interested in further considering Popkewitz' conclusion about the conditions necessary for long lasting contributions that linking development and policy research can make to academic disciplines (see Perspective section above), are studies that build upon the definition of "ambiguous" in Table 3. For example, a study at the Center for the Study of Writing (University of California, Berkeley) illustrated how educational policy analysts can create opportunities to analyze a large number of different reform polices aimed at inducing change by targeting major components of the instructional methods that are at the core of schooling (i.e., assignments, tests, grades, distribution criteria). Fillore and O'Connor (1986) combined knowledge gained from decades of policy research on schools as social systems and cultural systems resistant to change, with decades of language research discussions on how prevailing instructional assessment frameworks can be approached in terms of calls for socially responsible testing and assessment in a culturally pluralistic society. They demonstrated language ambiguity by illustrating the various ways students justify their choices for answers on reading and writing tests. Following Chafe's (1977), Florio's and Shultz' (1979) investigations of production and comprehension from the conceptual vantage point of language per se is ambiguous, Filmore and O'Conner's direction for policy studies of instructional assessment pointed out that there are differences in how individuals choose to make summarizing statements about the same topics. Among the choices identified to date are: (a) begin by summarizing an event and then giving details, or (b) build up details and then present the summary at the end (Filmore and O'Conner 1986). These discourse forms--or arrangements of sentences--are just another way of demonstrating and illustrating Freedle's (1980) compilation of findings across ethnographic studies of language use: first, that language forms are necessarily incomplete in specifying the full intentions of writers and speakers and so individuals choose schemas to help guide their selection for an answer; and, second, that language per se is ambiguous and so to comprehend an oral and written text individuals must necessarily initiate some interpretative frames to fill in needed information.

In other words, the concept of language per se is ambiguous is a schema/frame that is useful as a metacommunication tool for answering what must the next generation of policy analysts consider ... what can they strive to do in order to: (a) realize that what is not said is as important as what is said, (b) recognize the importance of ambiguity for creating choices or options, and (c) recognize the evolving nature of meaning, perhaps by considering Lakoff's image of a three-sided triangle.

Interpretative frames. The Table 3 glossary defines interpretative frames visible in and across functions of language. The definition is presented along with two questions to guide discovery of what is visible in and across functions of language that will be encountered in the social contexts policy researchers target for data gathering. The demonstrate that the concept of interpretative frames can be effectively used to develop inquiry statements for studying communication and change across the time frame of educational evaluation and policy analysis projects (cf. Freedle 1980; Hymes 1979; Gumperz & Cook-Gumperz 1982). This goal, of course, sounds quite ambitious. However, identifying and using language and policy concepts in research essentially involves making explicit the discourse resources we all have. For example, the chances of becoming explicitly aware of the idea of interpretative frame (and linking this concept to professional actions) seems particularly optimistic given the fact that educational policy researchers have had experience in using this language resource. One observation in research utilization literature is that different consequences can be anticipated from an author's use of academic, bureaucratic, or legal features of language (cf. Wallat & Piazza 1991). For many, these implicit understandings of the construction of meaning, and meaning in context, undergird their choice of specific discourse devices to persuade, inform, proclaim, or develop an argument. In other words, the conventions used to help make intent and meaning connections in the audiences' mind essentially adds up to constructing an interpretative frame.

This brief illustration of interpretative frame shows that intelligibility, or dealing with ambiguity, is not contingent upon finding an illusive set of "perfectly clear words" to connect intent and meaning. Rather the constraints and contingencies individuals take into account in constructing an interactive frame are the resources used to help disambiguate the authors' or speakers' intent (cf. Gumperz 1982a, 1982b; Wallat & Piazza 1991; Grimshaw 1987).

Schema/frames. Discourse theory and research address the consequences of background and other interactive experiences apparent during particular institutional routines in educational, health, and social services. Notions of knowledge structure and interpretation have been the object of study for a long time, and recent formulations of the concepts of frame and schema try to capture cognitive and socio-cultural dimensions of variations in and across contexts. For example, a recent attempt to answer how can we study comprehensive services for children and families built on the definition of schema/frames presented in Table 3. The idea of static schemas as personal understandings, relationships, values, goals, and interests held by individuals (in this case professionals and parents schemas about education and health, and what an educational and medical service should or can do to deal with educational or health matters) was combined with the idea of interpretative frames to investigate the ways in which activities conducted and actions taken in a comprehensive services identified and dealt with ambiguity and a mosaic of schemas (cf. Tannen & Wallat 1993).

Participant structure. There are several possible ways to arrange silence or articulation of multiple schemas and interpretative frames within the basic framework of verbal and nonverbal communication use in public life, including attempts to plan and deliver comprehensive services. Philips (1993) refers to these arrangements as "participant structure." These structural arrangements of discourse may fall into many different categories. In one type of participant structure one individual may address the entire group, or the talk may flow as if first - come - first - served reporting turns were being taken. Other participant structure arrangements include attention focused on one-to-one encounters between individuals, or attention focused on specific materials.

Studies of participant structures have identified an extensive set of context cues and strategies individuals use to constitute participant structures. At the same time such studies point to the use of a variety of meanings or interpretations for these context cues by both the researchers conducting the study and those participating in the project. Understandings of variant features of language use have led to new descriptions of the most common ways people verbally or nonverbally acknowledge and incorporate, or fail to incorporate or ratify, speakers utterances in and across participant structures (cf. Wallat 1984). Identifying and considering these cues and strategies can provide understandings of the links between the enactment of a particular policy and the participant structures in use during a project.

Construction of social norms. Giving recognition to identifying and considering the social and cognitive processes related to presenting information and group problem solving across participant structures centers on recognition of individuals as interpreters of their world(s) and as sources of influence on others. Based on this conceptualization, and in light of research on participant structures, the concept of construction of social and cognitive norms as process is proposed in discourse studies to contrast the view of norms as a discrete set of rules inculcated into passive participants (Wallat & Green 1982). As the definition in Table 3 points out, individuals must negotiate a complex system of arbitrary norms and rules within the forms of communication in today's social institutions. Norms and rules are arbitrary in the sense that definitions of the meanings of social situations, and situations in which not all rules and expectations are clearly stated, are evolving. Research analysis of social groups have added to the growing appreciation of norms of discourse as rules which can be modified, checked, suspended, terminated, and recommenced. These variable "correctness" rules for written discourse can be added to our understanding of variable "correctness" rules about speech: when to speak, how to speak (i.e., what gestures, movements, intonation, stress, and pitch features should be used), how to get a turn, how to digress from a topic appropriately, and to whom messages should be addressed. In order to establish and maintain social interaction, the participants must have agreed upon signals for beginning and ending a single social occasion. Therefore, the problem facing those concerned with identifying variable social norms or sociolinguistic rules for knowing when to speak is to capture the elements of a group's system of signaling coordination conventions including not only verbal statements but the participant's gestures in relation to objects or other persons in the environment (cf. Wallat & Green 1982, p. 101, 118).

Enactment of routines. One routine interpretation task that flows from attempts to construct social norms through changes in policy is the placement of children, youth, and adults in a host of human service programs (Szanton 1995). The way in which such everyday decisions are reached (or how information regarding these decisions is distributed in institutional settings), cannot be described simply by adding a few more factors to a comprehensive services model or to a model of social operation. Mehan's (1984) explication of the concept of enactment of routines defined in Table 3 suggests several possibilities for advancing our knowledge of an organization's functions in relation to changes in policy. The use of this concept in educational policy studies could help us consider how policy can be understood less as a set of "acts," "choices," and searches for "reasons" and "motives" and more as end results or consequences of variation in functions of language, interpretative frames and participants structures. Such study can stem from two interests. The first interest, as introduced above, is conceptual; to redefine production of policy reports as both an analytical and political opportunity for identifying certain social practices as symbolic of frames, participant structures and language functions that can signal to policy researchers possible interpretations held by individuals about policy objectives, choices and consequences. The second interest, also implied above, is practical. The next generation of policy researchers can be the benefactors of advances in understanding the advances in understanding the benefits and constraints of variations in our ideologies about language use as evident in the enactment of routines. The science of linguistics has been tied to investigating ideology through its concern with discovering and describing units of linguistic form; structures or patterns in which such forms are defined and situated; and the roles or functions that these units or forms serve in these structures (Fillmore 1985). The idea of a place for discussion about forms and functions is generally referred to as practical knowledge or metacommunication. The concept of enactment of routines, in combination with other concepts included in Table 3, fills in some of the detail of Popkewitz's (1991) and Conquergood's (1991) argument of the need to develop social epistemology as a legitimate art of inquiry, and Bailey's (1979) admonishment that educational policy researchers have simply got to learn something about rhetoric and style.

Style shifts. One basis for understanding participant structures and enactment of routines is semantics: How do people communicate and interpret meaning in everyday action and conversation? Each person's decisions about which communication strategies to apply across different situations results in her/his characteristic style (e.g., That style, then, is made up of a range on the continuum: the particular degree of camaraderie or deference in response to the situation, the people participating, the subject at hand, and so on). "Each person's notion of what strategy is appropriate to apply is influenced by a combination of family background and other interactive experience" (Tannen 1984, p. 14).

In terms of written discourse, we can arrive at new qualitative insights and new notions about our own and our colleagues' extensive communicative competence through the consideration of multiple definitions of style found in the literature, and the multiple approaches which have been created for understanding patterns established in a spoken or written text and the functions of the text.

Communicative competence. Studies of communicative competence have helped to widen the lens of both theory and research on what components of communication resources (besides grammatical diversity, stylistic resources and options, and interpretative processes) are essential in everyday life. It is generally accepted empirically, for example, that outsiders who enter a new demographic or professional scene have to ask "What's happening here?" and may have to adapt or shift linguistic codes in order for their policy project to be considered appropriate. Gumperz (1982a; 1982b) has discussed the theoretical and practical aspects of entry and access as "establishing a successfully negotiated frame of reference." He illustrates this concept through reports of what is going on across multiple bureaucratic, business, and judicial contexts. In each of these policy study related investigations, the researchers efforts at locating the overall institutional frame of reference began with time and effort spent on reaching agreement with informants on what activity both parties believed was being enacted. As Gumperz points out, outsiders who enter a new demographic or professional scene have generally learned a new code at the level of lexicon, and this knowledge will be sufficient for the instrumental contacts that fill up much of their project working day. But situations of persuasion, involving the ability to explain, describe, or narrate, are often difficult to manage. Here breakdowns tend to lead to mutual stereotyping and pejorative evaluations. To be sure not all problems of inter [group] contact are communicative in nature. Economic factors, differences in goals and aspirations, as well as other historical and cultural factors may be at issue. (Gumperz 1982c, p. 331)

But there is reason to suspect that a significant number of communication breakdowns may be due to the failure to establish a successfully negotiated frame of interpretation. This neglect of variation in individual and social styles has been identified in studies of: participant structure differences across contexts (e.g., Green 1983; Cazden 1986); interactive frames and knowledge structure schemas across contexts (e.g., Tannen & Wallat 1993); and characteristics of dozens of visible information cues (or what sociolinguists refer to as contextualization cues), that have been illuminated in a range of communicative contexts (cf., Wallat 1984). According to studies of language in public life designed and conducted by Professor John Gumperz (1982a; 1982b), breakdowns can be traced to a lack of time and effort spent by the researchers and informants on reaching agreement on what activity is being enacted and how it is conducted.

> Distinctions among such activities as chatting, discussing, and lecturing exist in all cultures, but each culture has its own constraints not only on content but also on the ways in which particular activities are carried out and signaled. Even within a culture, what one person would identify as 'lecturing' another might interpret as 'chatting with one's child.'.... Since speech activities are realized in action and since their identification is a

> function of ethnic and communicative background, special problems arise in modern society where people have widely varying communicative and cultural backgrounds. How can we be certain that our interpretation of what activity is being signaled is the same as the activity that the interlocutor has in mind, if our communicative backgrounds are not identical? (Gumperz 1982a, p. 166-167)

SECTION TWO SUMMARY

The illustrative definitions included in Table 3 provide an overview of basic concepts researchers have identified, defined, and used in applying sociolinguistics, discourse analysis, conversational analysis and ethnography of communication methods in studies of educational issues (e.g., Briggs 1986; Erickson 1975; Gumperz and Cook-Gumperz 1982; Gumperz & Hymes 1972; Mishler 1986). More importantly, perhaps, Table 3 illustrates social constructs that provide a rationale and framework for investigating how oral and written discourses actually work across institutional settings. This set of key concepts may also begin to help those interested in educational policy issues become sensitive to observing, rather than taking for granted, communicative actions throughout a project. They are offered not for the sake of detailed analysis of oral and written phenomena in the sites and activities encompassed by a project, but because they may shed light upon the perennial problem of accounting for the variety of ways in which oral and written communicative behavior is organized in interaction in different situations.

The final section provides an elaboration of the idea that researchers and policy makers often have different "definitions of a situation." Using the concepts presented in Table 3 we discuss the particular functions served by two approaches to early childhood education and policy analysis. The choice of these two approaches was based on the data presented in Table 1, that is to consider two approaches which appeared the most often in EEPA during the fifteen year review period reported in here.

USING A COMMUNICATIVE FRAMEWORK: AN APPLICATION ACROSS TWO EXAMPLES OF POLICY REPORT FUNCTIONS

Throughout the first fifteen years of EEPA's publication, contributors have suggested a need to develop a conceptual framework that takes into account variations in perspectives and meanings held by clients and audience. (cf. Bolland & Bolland 1984; Borich 1983; Felter 1986; Hayman, Rayder, Stenner & Madley 1979; Hood 1985; House 1979a, 1979b; Kelly 1980; Kenny 1982; Lynch 1983; Maxwell 1984; Page 1979; Sadler 1985; Shapiro 1985; Stake 1981; Stufflebeam & Webster 1980; Williams 1986; Wortman 1982). We turn now to showing how the communication concepts discussed can be specifically applied to the various policy approaches identified in EEPA. We use two examples, accountability and case study.

Descriptions of ACCOUNTABILITY projects range from cost-benefit studies for monitoring the use of funds (Kean & Scanlon 1979:

House 1979b; Warfield 1994); to projects aimed at meeting technical requirements such as computing percentages and setting cutoff scores (Nations 1982; Smith 1982); to developing systems for keeping records of attendance, expenditures, and test scores in order to disseminate evaluation results, identify exemplary programs, or find the simplest way to meet state and federal requirements (Barnes & Ginsburg 1979; David 1981; Dougherty 1979; Nations 1982; venderPloeg 1982); to projects concerned with providing constituents with an accurate accounting of results (Stufflebeam & Webster 1980), or a demonstration of "the responsiveness and the political responsibility of public school institutions" (Cohen, 1979 p.59). Additional indicators that accountability is an evolving social invention in evaluation approaches include its role in justifying action(s) or serving as the raison d'etre for protracted negotiation at Congressional and agency levels (Benveniste 1985; Cross 1979; House 1979b; Stonehill and Grover 1983).

When asked why this burgeoning of educational evaluation has occurred since 1970, Gorwin and Green (1980) answered that the phenomenon arose and developed itself out of a process of extended legislative authority that required something called an evaluation. In contrast (i.e., in an academic rather than a political interpretive frame), Stufflebeam and Webster (1980), answered the same question by suggesting that the extension of evaluation into accountability studies was the result of the pioneering efforts of Lessinger in his 1970 book *Every Kid A Winner: Accountability in Education.*

Development and policy research readers have the opportunity to make up their own minds as to whether the developments in accountability functions are due to specific legislative acts or specific academic accomplishments, or whether they result from countless face-to-face encounters involving evaluation at Congressional, Federal agency, State agency, and local levels. At least eight policy researchers who report on accountability developments in early childhood provide close up views of how accountability provisions are created and modified across these levels. Each provides a unique contextual interpretation of what actually goes on across policy arenas (Barnes and Ginsberg 1979; Cross 1979; House 1979a; Marshall, Mitchell & Wirt 1986; Sacken & Medina 1990; Slavin & Madden 1991; Stonehill and Grover 1983; venderPloeg 1982; Wells & Peterson 1992; Wirt, Mitchell & Marshall 1988). Overall, when reading the discussion of procedures in accountability articles in volumes one through sixteen, one discovers new understandings of the many facets of communicative processes that enhance or diminish compliance and the impact of law. As Kaufman (1984) has noted, those who are attempting to deal with questions of law must contend with all the drama, confusion, failure, and achievement that constitutes interpreting the meaning and purposes of a particular statute.

Several concepts from communication theory and research may be particularly useful for educational researchers in dealing with the paradox of law - that law never is, but is always about to be realized when embodied in a judgment. The concept of INTERPRETATIVE FRAMES is particularly useful since its purpose is to help explain variations in participants' judgments in the interpretation of discourses. This concept also helps focus such accountability research tasks as how to deal with multiple interpretation processes that individuals may use to frame their understanding of the relation of law to action. Not only would it be possible to consider new indicators of policy terms such as "impact", but the adoption of an analytical stance that includes interpretative frame and STYLE OPTIONS/ SHIFTS would help researchers account for the language related tasks all accountability projects now only implicitly are able to include in project cost estimates: (a) constructing an understanding of the contours and contexts in which a legal point arose, (b) constructing an understanding of the complex interplay of language and intent, and (c) constructing an understanding of the complex strands of thought that individuals use in relation to law in action (cf. Wallat 1987).

Concepts such as FRAMES/SCHEMA could add a discourse perspective starting point to educational evaluation accountability approaches to determining whether particular policies were in fact accomplishing specific objectives of a law. For example, mandates for collaboration have been included in policy since the mid-1980's. The contribution of a communicative framework to determining the consequences of collaboration on the forms of communication which take place between providers, and between providers and clients, has been demonstrated. Research that includes the concepts of "frames" to refer to the anthropological/sociological notion of interactive frames of interpretation, and "schema" to refer to the psychological/artificial intelligence notion of knowledge schemas, have been included in the design of studies of family and child development centers (Tannen & Wallat 1993). Findings from such studies suggest that overlapping, competing, and possibly conflicting frames are inherent in the structure of parent/professional interaction in particular and communication in general. These are forces at work that can at times create problems in the best of all possible educational, health, or social worlds.

Perhaps the clearest example of how educational policy study discourse functions have been created, modified, expanded, or transformed during the past fifteen years is the variety of meanings of the term accountability legislation. In 1980, Kirst and Jung pointed out that one largely overlooked issue in policy research is knowing "how to extract knowledge from the information we already have" (p. 31). In their example of how attention can be focused in policy work, knowledge is equated with a survey of the language of policy documents, including relevant statutes and subsequent amendments, original regulations and subsequent modifications, audit checklists on compliance reports, legislature hearings, and public documents that detail key sequences of events. Such documentary analysis provides a means of developing a manageable data scope for tracing changes in a legal framework over time. Some of the reasons why accountability evaluation enables tracing changes in initial legislation and affiliated legal documents, or enables detecting shifts in objectives and priorities, are simply due to the FUNCTIONS OF legal language; legal writers attempts to deal with the AMBIGUOUS nature of language, and the CONSTRUCTION OF NORMS for the STYLE of judicial discourse. As Kirst and Jung point out, statutes often are formulated with deliberately obfuscated language to broaden political support for legislation:

> Winning coalitions are often held together by the adhesive of ambiguous language which successfully masks unresolved differences among competing interest groups and legislators. Statutes are as heavily laced with symbolic rhetoric as they are replete with allocative formulae and regulatory prescriptions.... Therefore, in order to detect shifts in objectives and priorities, it is also necessary to closely analyze those documents which, in essence, operationalize the symbolic import characteristics of most statutes. These include regulations, mandatory and explanatory criteria, guidelines, technical assistance packages, audit check lists, application forms, evaluation mechanisms, and complaint resolution processes. (Kirst & Jung 1980,p.29)

Development and policy researchers who need to analyze legal documents are reminded that the functions served by legal style features (use of archaic words, infusion of ordinary words such as "real property" with specific legal meaning, use of embedded clauses) require one to shift gears in the frame of reference or schema they use in reading. The function of legal language is "to allow one expert to register information for scrutiny by another" in contrast to the function of other texts one may read (such as plans for an organized activity for children and their families) (Crystal & Davy 1969, p 193).

CASE STUDY projects reported in EEPA also provide clear examples of how an approach is affected by new creations, modifications, expansions and transformations of the meanings of its functions. Researchers such as Bissell (1979), Monti (1979), Mercurio (1979), Kirst and Jung (1980), Porter (1983), and Mazzoni and Clugston (1987), have formulated arguments regarding the usefulness of identifying how variable strategies of administration enable a new policy to be introduced into a school district's routines while change is avoided; how field study techniques used in case studies could provide a sense of expanding dimensions of citizen involvement mandated in a variety of compensatory education programs; how case studies can be useful for identifying the language of state-wide reform efforts and policy enactment variables and mapping relations among these variables; and, how use of an interactionist perspective in doing case studies would result in documentation of different interpretations of an education innovation rather than reporting one "correct" interpretation or description.

Other researchers who have written about case study have tried to detail some common communicative elements in data collection procedures (e.g., primarily interview based), some common problems to expect in any attempt to achieve a comparative analysis of situations or sites (Alkin and Daillak 1979; Sadler 1985), some general format/style categories to use in writing up case studies (Barnette 1983) and some possibilities for how a case study evaluation of evaluations conducted over a ten or thirteen year period could pull together alternative views and identify new indicators of impact without the cost of collecting new data (St. Pierre 1982; Kirst and Jung 1980).

A central issue of the case study is what it means to describe a process. It is almost a platitude to say that the meaning of a process depends upon its context, or whatever factors or curriculum content in the learning environment the research directs attention to. Researchers concerned with understanding variability across sites and cases all agree that determining processes, or what is meant at any one point in time, depends upon finding empirical methods capable of determining the extent to which underlying knowledge and behavioral norms are shared (cf. Gumperz 1982a). Their experiences with discourse analysis methodologies suggest that "all participants must be able to fit individual contributions into some overall theme roughly corresponding to a culturally identifiable activity, or a combination of these, and agree on relevant behavioral norms" (Gumperz 1982a, p. 163). This idea, or the concepts of PARTICIPANT STRUCTURE, CONSTRUCTION OF SOCIAL NORMS, and ENACTMENT OF ROUTINES seems particularly useful for helping focus case study researchers on the value of drawing on comparative data on the ordinary parts of a school's work day: on what school personnel do, how much they do, and how they produce forms of social organization and thereby construct such recognizable characteristics as not solving a problem, marking something as problematic, or bringing to public notice what everyone knew anyway. In other words, with adoption of a communicative framework, case study policy researchers can contribute to understandings of what adopting an educational schema/frame actually means as an observable, interactional feature of daily public life (cf. Anderson, Hughes, and Sharrock 1987).

These researchers' work suggests that an individual enacting a case study discourse function essentially follows a communicative policy-analysis model by focusing detailed attention on a limited number of policy questions that have salience across locations and settings. Among the questions early childhood evaluators and policy researchers have addressed are : How does the community and staff react to and/or perceive a policy such as the California reforms which focused on elementary mathematics? (Ball 1990; Cohen 1990; Peterson 1990). Case studies of the California attempt to create substantial change in instruction to foster deeper understanding of mathematics and to improve students' capacity to reason mathematically revealed the complexity of teachers interpretative frames on mathematics learning. They also exposed the tangled influences of policy and the difficulties inherent in communicating the California Mathematics Curriculum Framework policy.

SECTION THREE SUMMARY

Those who attempt to practice using concepts such as FUNCTIONS OF LANGUAGE, LANGUAGE PER SE IS AMBIGUOUS, INTERPRETATIVE FRAMES, SCHEMA/ FRAMES, PARTICIPANT STRUCTURES, CONSTRUCTION OF SOCIAL NORMS, ENACTMENT OF ROUTINES, STYLE OPTIONS/ SHIFTS will find that the case studies of the California reforms as well as recent case studies of Title 1/Chapter One (C. de Baca, Rinaldi, Billig, & Kinneson 1991; Winfield 1991) lend themselves to thinking about the multiple examples of the language resources individuals use in public life. Use of these concepts may also bring to mind the value of new evaluation and policy analysis which addresses questions such as: What sort of preparations can be made in a school under legal mandate? What does the policy look like in practice across multiple interactions? What type of curriculum is in use, and what, if any, changes can be made in it? What type of policy do parents and teachers really want? How much time are they willing to give to deal with consequences which can be traced to the use of everyday language?

As demonstrated in the accountability and case study studies published in the first fifteen years of publication of one educational evaluation and policy studies journal, multiple references to communication activities are made within these approaches. The illustration of language and policy concepts in the review of the discourse function of accountability and case study evaluation and policy analysis suggest that development and policy researchers can act on advice to learn more about participation processes; participants' points of view and definitions of situation; and ambiguous features of language in policy documents, in evolving social structures, and in different value perspectives (cf. Cousins and Earl 1992). In effect these evaluation and policy research concerns, and the set of concepts included in the communicative framework presented, describe a venue that can be undertaken through further consideration of the relationship between language and policy by a new generation of "methodologically multilingual" professions.

FINAL THOUGHTS FOR PUTTING IT ALL TOGETHER: THE CASE FOR LINKING DEVELOPMENT AND POLICY RESEARCH

As illustrated in propositions such as "a reinvestment cycle," the issues at stake in child and family policy go well beyond understanding established technical criteria for defining the kind of academic information which can be provided and the cognitive features measured (cf. Oakes 1989). To persuade the U.S. public that infancy, preschool, early education and other services for children and families are effective, professionals have been called upon since the early 1980's to expand their research into the complexities of linking development and public policy issues (Datta 1983, p. 144).

The point of this article has been that the major challenge for the next generation of development and policy researchers is to be well versed in the social-communicative functions served by available approaches in educational evaluation and policy analysis, methodologically multilingual in using the forms of multiple approaches with consumer and policy audiences, and capable of articulating variable interpretations of educational outcomes.

To help set solutions to this challenge, this article suggested how studies of language and policy can help policy researchers do two things: (a) consider evaluation and policy analysis approaches as social inventions that serve a range of communicative functions, and (b) consider the link between child and family development research and policy analysis and discourse literature through use of a set of communicative concepts that provide the foundation for meeting the demand for a participatory policy discourse, including insiders accounts of events, objects and actions.

End Notes

1. See Conquergood (1991) for review of contributions to explicating these concepts which are included in "boundary" perspectives across studies .

2. Many attempts have been made to classify research questions into propositions defined as the varieties of linkage between determinants and results, e.g., ZETTERBERG, H.L. (1965) On theory and verification in sociology. (Totowa, NJ: The Bedminster Press), or, following Aristotle, as classes of predicates that might be formed about educational research and educational policy research in particular (e.g., DILLON, J. T. (1984) "The classification of research questions". Review of Educational Research, 54 (3), pp. 327 - 361; SMITH, N. L. & MUKHERJEE, P. (1994) "Classifying research questions addresses in evaluation studies". Educational Evaluation and Policy Analysis, 16 (2), pp. 223 - 230). For examples of propositions within a language perspective see : SILVERMAN, D. & TORODE, B. (1980) The material word: Some theories of language and its limits. (Boston : Routledge & Kegan Paul); GOODENOUGH, W. H. (1990) "Evolution of the human capacity for beliefs". American Anthropologist, 92 (3), pp. 597 - 612.

3. Identification of benefits such as increases in school success, employability, and self-esteem; returns of $4.75 in lowering costs of special education for $1.00 investment in preschool; decreases in the $3000 cost of repeating a grade; and savings of $1560 per disabled pupil because of early education intervention, have been associated with Congressional members endorsement of programs designed for children from birth through the first few years of elementary school as "Opportunities for Success" (Select Committee on Children, Youth and Families 1985, Washington, D.C.).

References

Alkin, M. C., & Dallak, R. H. (1979). A study of evaluation utilization. *Educational Evaluation and Policy Analysis,* 1 (4), 41-49.

Anderson, R. J., Hugher, J. A., & Sharrock, W. W. (1987). Executive problem solving : Some material and initial observations. *Social Psychology Quarterly,* 50(2), 143-159.

Atkinson, P. (1990). *The ethnographic imagination: textual construction of reality.* New York: Routledge, Chapman and Hall, Inc.

Bailey, S. T. (1979). Current educational policy analysis: An insight [Interview]. *Educational Evaluation and Policy Analysis*, 1 (3), 95- 100.

Ball,L. D. (1990). Reflections and deflections of policy: The case of Carol Turner. *Educational Evaluation and Policy Analysis* , 12, (3), 247 - 260.

Barnes, R. E., & Ginsbury, A. L. (1979). Relevance of the RMC Models for Title I policy concerns. *Educational Evaluation and Policy Analysis*, 1(2), 7-14.

Barnett, W. S., Fredee, E. C., Mobasher, H., & Mohr, P. (1988). The efficacy of public preschool programs and the relationship of program quality to efficacy. *Educational Evaluation and Policy Analysis,* 10(1), 37-49.

Barnette, J. J. (1983). Naturalistic study of project interagency linkages: The linkage case study. *Educational Evaluation and Policy Analysis*, 5(4), 473-483.

Becker, A. L. (1988). Language in particular: A lecture. In D. Tannen (Ed.), *Linguistics in context. connecting observation and understanding* (pp. 17- 35). Norwood, NJ: Ablex.

Benveniste, G. (1985). The design of school accountability systems. *Educational Evaluation and Policy Analysis,* 7(3), 261-279.

Benveniste, G. (1991). [Review of Evaluation as policymaking: Introducing evaluation into a national decentralized system]. *Educational Evaluation and Policy Analysis,* 13, (3), 322-323.

Birman, B. F. (1981). Problems of overlap between Title I and P.L. 94-142: Implications for the federal role in education. *Educational Evaluation and Policy Analysis*, 33(3), 5- 20.

Bissell, J.S. (1979). Use of educational evaluation and policy studies by the California legislature. *Educational Evaluation and Policy Analysis*, 1 (3), 29-37.

Bolland, J. M., & Bolland, K. A. (1984). Program evaluation and policy analysis: Toward a new synthesis. *Educational Evaluation and Policy Analysis*, 6(4), 333-340.

Borich G. G. (1983). Evaluation models: A question of purpose not terminology. *Educational Evaluation and Policy Analysis*, 5(1), 61- 64.

Briggs, C.L. (1986). *Learning how to ask: A sociolinguistic appraisal of the role of the interview in social science research.* New York: Cambridge University Press.

Bruner J. S. (1966). Education as social invention. In J. S. Bruner, *Toward a theory of instruction* (pp.22-38). New York: W.W. Norton & Co., Inc.

Caracelli, V. J., & Greene, J. C. (1993). Data analysis strategies for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 15, (2), 195-207.

Cazden, C. B. (1986). Classroom discourse. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 432-463). New York: Macmillan.

Chafe, W. (1977). Creativity in verbalization and its implication for the nature of stored knowledge. In R.Freedle (Ed.), *Discourse production and comprehension* (pp. 41-56). Norwood, NJ: Ablex.

Cicourel, A. V. (1987). The interpenetration of communicative contexts. *Social Psychology Quarterly*, 50, (2), 217-226.

Cicourel, A. V. (1994). Theoretical and methodological suggestions for using discourse to recreate aspects of social structure. In A. D. Grimshaw, P.T. Burke, J. Cook-Gumperz, S. Feld, C. J. Fillmore, L. W. Fillmore, J. J. Gumperz, M.A.K. Halliday, R. Hasan, & D. Jenness, *What's going on here? Complementary studies of professional talk.* (pp. 61 - 89). Norwood, N.J.: Ablex Pub. Corp.

Cohen, D. K. (1979). Educational policy-making: Private influence and citizen participation. *Educational Evaluation and Policy Analysis,* 1(5), 59-65.

Cohen, D. K. (1990). A revolution in one classroom: The case of Mrs. Oublier. *EEPA,* 12, (3), 311-330.

Cohen, D. K. & Garet, M.S. (1975). Reforming educational policy work with applied social research. *Harvard Educational Review*, 45 (1), 17-43.

Conquergood, D. (1991). Rethinking ethnography: Towards a critical cultural politics. *Communication Monographs*, 58, 179-194.

Cooley, W. W., & Leinhardt, G. (1980). The instructional dimensions study. *Educational Evaluation and Policy Analysis*, 2 (1), 7-26.

Cousins J. B., & Earl, L, M. (1992). The case for participatory evaluation. *Educational Evaluation and Policy Analysis*, 14, (4), 397- 418.

Crain, R. L., & Mahard, R. E. (1979). Research on school desegregation and achievement: How to combine scholarship and policy relevance. *Educational Evaluation and Policy Analysis*, 1(4), 5-15.

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist*, 2, 116-126.

Cross, C. T. (1979). Title 1 evaluation: A case study of Congressional frustration. *Educational Evaluation and Policy Analysis*, 1(2), 15- 21.

Crystal, D. (1980). *A first dictionary of linguistics and phonetics.* Boulder, CO: Westview Press.

Crystal, D. (1987). *The Cambridge encyclopedia of language*. New York:

Cambridge University Press.

Crystal, D., & Davy, D. (1969). *Investigating english style.* Bloomington, IN: Indiana University Press..pa

Darlington, R. B. (1981). The Consortium for longitudinal studies. *Educational Evaluation and Policy Analysis*, 3(6), 37- 45.

Datta, L. E. (1983). Epilogue: We never promised you a rose garden, but one may have grown anyway. Consortium for Longitudinal Studies, *As the twig is bent ... Lasting effects of preschool programs* (pp. 467 - 479). Hillsdale, NJ: Lawrence Erlbaum Associates.

Datta, L. E. (1982). Strange bedfellows: The politics of qualitative methods. *American Behavioral Scientist*, 26 (1), 133-144.

David, J.L. (1981). Local uses of title 1 evaluations. *Educational Evaluation and Policy Analysis*, 3 (1), 27-39.

Debaca, M.R.C., Rinaldt, C., Billig, S. H., Kinnison, B. M. (1991). Santo Domingo School: A rural schoolwide project success. *Educational Evaluation and Policy Analysis*, 13, (4), 369- 380.

Demauro, T. (1967). *Ludwig Wittgenstein: His place in the development of semantics.* Netherlands: D. Reidel.

Destefano, L. (1992). Evaluating effectiveness: A comparison of federal expectations and local capabilities for evaluation among federally funded model demonstration programs. *Educational Evaluation and Policy Analysis*, (14,) 2, 157- 168.

Dore, J., & McDermott, R. P. (1982). Linguistic indeterminacy and social context in utterance interpretation. *Language*, 58 (2), 374-398.

Door-Bremme, D.W. (1985). Ethnographic evaluation: A theory and method. *Educational Evaluation and Policy Analysis*, 7 (1), 65- 83.

Dougharty, L. A. (1979). A policy analysis of the critical mass concept. *Educational Evaluation and Policy Analysis*, 1(3), 5-18.

Easton, P. A. (1989). Structuring learning environment: Lessons from the organization of post- literacy programs. *International Review of Education*, 35, 423-444.

Eisner, E. W. (1979). Toward a methodology of naturalistic inquiry in educational evaluation. *Educational Evaluation and Policy Analysis*, 1(2), 73-74.

Ellwein M.C., Walsch, D. J., Eads, G. M., & Miller, A. (1991). Using tests to route kindergarten students: The snarled intersection of psychometrics, policy and practice. *Educational Evaluation and Policy*

Analysis, 13, (2), 159-175.

Erickson, F. (1975). Afterwords. In A. Kendon, R.M. Harris, & M.R. Key (Eds.), *Organization of behavior in face-to-face interaction* (pp. 483-485). Chicago, IL: Aldine (distributors for Mouton Pub., The Hague).

Erickson, F. (1978, October). *Timing and context in everyday discourse: Implications for the study of referential meaning.* Paper presented at the Conference on Children's Oral Communication Skills, Wisconsin Research and Development Center for Individualized Schooling, University of Wisconsin at Madison.

Erickson, F., & Shultz, J. (1982). *The counselor as gatekeeper: Social interaction interviews.* New York: Academic Press.

Escobar, C. M., Barnett, W. S., & Keith, J. E. (1988). A contingent validation approach to measuring the benefits of preschool education. *Educational Evaluation and Policy Analysis,* 10(1), 13-22.

Fetler, M. (1986). Accountability in California public schools. *Educational Evaluation and Policy Analysis*, 8(1), 31-44.

Filmore, C. (1985). Linguistics as a tool for discourse analysis. In T. A. van Dijk (Ed.), *Handbook of discourse analysis, Vol. 1* (pp. 11 - 39). New York: Academic Press.

Filmore, C., & O'Conner, M.C. (1986, April). *Findings of the Berkeley reading project.* Paper presented at the Conference on Research Perspectives on Language, Literacy and Schooling, University of California, Berkeley, Center for the Study of Writing.

Florio, D. H. (1979). Education and the politics of austerity. *Educational Evaluation and Policy Analysis*, 1(2), 75-82.

Florio, S. & Shultz, J. (1979). Social competence at home and school. *Theory into practice*, 18 (4), 234 - 243.

Freedle, R. (1980, April). *Interaction of language use with ethnography and cognition.* Paper presented at the meeting of the American Educational Research Association, Boston, MA. Available from author, Educational Testing Service, Princeton, NJ.

Gersten, R. M., Carnine, D. W., & Williams, P.B. (1982). Measuring implementation of a structured educational model in an urban school district: An observational approach. *Educational Evaluation and Policy Analysis*, 4(1), 67-79.

Gowin, D.B., & Green, T.F. (1980). Two philosophers view evaluation. [Interview with N.L. Smith, director of the Northwest Regional Educational Laboratory's Research on Evaluation Program]. *Educational Evaluation and Policy Analysis*, 2(2), 67-70.

Green, J. C., Caracelli, V.J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 11 (3), 255-274.

Green, J. L. (1983). Exploring classroom discourse: Linguistic perspectives on teaching-learning processes. *Educational Psychologist*, 18(3), 180-199.

Grimshaw, A. D. (1987). Disambiguating discourse: Members' skill and analysts' problem. *Social Psychology Quarterly*, 50 (2), 186-204.

Gumperz, J.J. (1982a). *Discourse strategies.* New York: Cambridge University Press.

Gumperz, J.J. (Ed.) (1982b). *Language and social identity.* New York: Cambridge University Press.

Gumperz, J.J. (1982c). The linguistic basis of communicative competence. In D. Tannen (Ed.), *Analyzing discourse: Text and talk* (pp. 323-334). Washington, DC: Georgetown University Press.

Gumperz, J.J., & Cook-Gumperz, J. (1982). Introduction: Language and the communication of social identity. In J.J. Gumperz (Ed.), *Language and social identity* (pp. 1-21). New York: Cambridge University Press.

Gumperz, J.J., & Hymes, D. (1972). *Directions in sociolinguistics: The ethnography of communication.* New York: Holt, Rinehart & Winston.

Halliday, M.A.K. (1973). *Explorations in the functions of language.* London: Edward Arnold.

Handleman, D. (1975). Domains of definition in interaction. In A. Kendon, R. M. Harris, & M. R. Key (Eds.), *Organization of behavior in face-to-face interaction.* (pp. 389 -414). Chicago,IL.: Aldine {distributors for Mouton Publishers, The Hague)

Hayman, J., Rayder, N., Stenner, A.J., & Madley, D.L. (1979). On aggregation, generalization, and utility in educational evaluation. *Educational Evaluation and Policy Analysis*, 1(4), 31-39.

Heath, R. W. & Brandon, P.R. (1982). An alternative approach to the evaluation of educational and social programs. *Educational Evaluation and Policy Analysis*, 4(4), 477-486.

Hood, P. D. (1985). Evaluation research and program evaluation: The view of a program director. *Educational Evaluation and Policy Analysis*, 7(3), 238-242.

House, E.R. (1979a). Coherence and credibility: The aesthetics of evaluation. *Educational Evaluation and Policy Analysis*, 1, (5), 5- 17.

House, E.R. (1979b). The objectivity, fairness and justice of federal evaluation policy as reflected in the Follow Through evaluation, *Educational Evaluation and Policy Analysis*,1 (1), 28-42.

Huston, A.C., MCLoyd, V.C., & Coll, C.G. (1994). Children and poverty: Issues in contemporary research. *Child Development*, 65, (2), 275-282.

Hymes, D. (1994). Discussion [*Theoretical and methodological suggestions for using discourse to recreate aspects of social structure*]. In A.D. Grimshaw, et al (Eds.) *What's going on here? Complementary studies of professional talk* (pp. 89 - 94). Norwood, N. J. : Ablex Pub. Corp.

Hymes, D. (1974). *Foundations in sociolinguistics: an ethnographic approach.* Philadelphia, PA.: University of Pennsylvania Press.

Hymes, D. (1972). Introduction. In C.B. Cazden, V.P. John, & D. Hymes (Eds.), *Functions of language in the classroom* (pp. i-ivii). New York: Teachers College Press.

Hymes, D. (1979). Language in education: Forward to fundamentals. In O.K. Garnica & M.L. King (Eds.), *Language, children and society* (pp. 1-20). Elmsford, NY: Pergamon Press.

Jordan, B. (1981). An EEPA interview with Barbara Jordon. *Educational Evaluation and Policy Analysis*, 3(6), 79-82.

Kaufman, I. R. (1984). The anatomy of decisionmaking. *Fordam Law Review*, 53, 1-22.

Kean, M. H., & Scanlon, R. G. (1979). The achieve - out phenomenon in Title I. *Educational Evaluation and Policy Analysis,* 1(4), 65-70.

Kelly, E. F. (1980). Evaluation as persuasion: A practical argument. *Educational Evaluation and Policy Analysis*, 2(5), 35-38.

Kenny, D. A. (1982). Book review: Evaluating with validity. *Educational Evaluation and Policy Analysis*, 4(1), 121-122.

Kirst, M. W. (1980). [An EEPA interview with Michael W. Kirst]. *Educational Evaluation and Policy Analysis*, 2(1), 77- 83.

Kirst, M. W., & Jung, R. (1980). The utility of a longitudinal approach in assessing implementation: A thirteen year view of Title I, ESEA. *Educational Evaluation and Policy Analysis*, 2(5), 17-34.

Klein, S. S., Gold, N., & Stalford, C. (1986). The convening process: A new technique for applying knowledge to practice. *Educational Evaluation and Policy Analysis*, 8(2), 189-204.

Koppelman, K. L. (1979). The explication model: An anthropological approach to program evaluation. *Educational Evaluation and Policy Analysis*, 1(4), 59-64.

Labelle, T.J. , Moll, L. C., & Weisner, T. S. (1979). Context-based educational evaluation: A participant research strategy. *Educational Evaluation and Policy Analysis,* 1(3), 85-93.

Lakoff, R. T. (1990). *Talking Power: The politics of language in our lives.* New York : Basic Books.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by.* Chicago, IL.: The University of Chicago Press.

Lazar, I. (1983). Discussion and implications of the findings. Consortium for Longitudinal Studies, *As the twig is bent... Lasting effects of preschool programs* (pp. 461 - 466). Hillsdale, NJ: Lawrence Erlbaum Associates.

Le Mahieu, P. G. (1984). The effects on achievement and instructional content of a program of student monitoring through frequent testing. *Educational Evaluation and Policy Analysis*, 6(2), 175-188.

Linn, R. L. (1979). Validity of inferences based on the proposed Title I evaluation models. *Educational Evaluation and Policy Analysis*, 1(2), 23-32.

Lunch, K. B. (1983). Qualitative and quantitative: Two terms in search of a meaning. *Educational Evaluation and Policy Analysis*, 5(4), 461-464.

MacPhee, D., Kreutzer, J. C., & Fritz, J.J. (1994). Infusing a diversity perspective into human development. *Child Development*, 65 (2), 699-715.

Marshall, C., Mitchell, D., & Wirt, F. (1986). The context of state-level policy formulation. *Educational Evaluation and Policy Analysis*, 8 (4), 347-378.

Marshall, C. & Rossman, G.B. (1989). *Designing qualitative research.* Newbury Park, CA.: Sage Publications.

Maxwell, G. S. (1984). A rating scale for assessing the quality of responsive / illuminative evaluation. *Educational Evaluation and Policy Analysis,* 6(2), 131-138.

Mazzoni, T. L., & Clugston, R. M. (1987). Big business as a policy innovator in state school reform: A Minnesota case. *Educational Evaluation and Policy Analysis*, 94(4), 312- 324.

McLaughlin, M. W. (1987). Learning from experience: Lessons from policy implementation. *Educational Evaluation and Policy Analysis*, 9 (2), 171-178.

Mehan, H. (1984). Institutional decision-making. In B. Rogoff & J. Lave (Eds.), *Everyday cognition: Its development in social context* (pp. 41-66). Cambridge, MA: Harvard University Press.

Meisels, S. J. (1985). A functional analysis of the evolution of public policy for handicapped young children. *Educational Evaluation and Policy Analysis,* 7(2), 115- 126.

Mercurio, J. A. (1979). Community involvement in cooperative decision making: Some lessons learned. *Educational Evaluation and Policy Analysis*, 1(6), 37- 46.

Mishler, E.G. (1986). *Research interviewing: context and narrative.* Cambridge, MA: Harvard University Press.

Monti, D. J. (1979). Administrative discrimination in the implementation of desegregation policies. *Educational Evaluation and Policy Analysis*, 1(4), 17-25.

Murphy, J. T. (1980). The state role in education: Past research and future directions. *Educational Evaluation and Policy Analysis,* 2(4), 39-51.

Nations, R. (1982). School district problems in evaluation of the state funded compensatory education programs.: A view from Florida. *Educational Evaluation and Policy Analysis*, 4(1), 21-31.

Nielsen, H. B. (1995). Seductive texts with serious intentions. *Educational Researcher*, 24 (1), 4-12.

Oakes, J. (1986). *Educational indicators: A guide for policymakers.* Santa Monica, CA.: The Rand Corporation.

Odom, S. L., & Fewell, R. R. (1983). Program evaluation in early childhood special education: A meta evaluation. *Educational Evaluation and Policy Analysis*, 5(4), 445-460.

Osterlind, S. J. (1979). What makes a school evaluation work? *Educational Evaluation and Policy Analysis*, 1 (1), 48-51.

Page, E. P. (1979). Counterpoint: More Objective! *Educational Evaluation and Policy Analysis*, 1(1), 45-66.

Peterson, P. L. (1990). Doing more of the same: Cathy Swift. *Educational Evaluation and Policy Analysis*, 12, (3), 261- 280.

Philips, S.U. (1993). *The invisible culture.* Prospect Heights, IL.: Waveland Press.

Popham, W. J. (1979). Statement of EEPA editorial policy. *Educational Evaluation and Policy Analysis*, 1(1), 1.

Popkewitz, T. S. (1991). *A political sociology of educational reform.* New York: Teachers College Press.

Porter, P. (1980). Policy perspective on the study of educational innovations. *Educational Evaluation and Policy Analysis*, 2(4), 73- 84.

Quie A. H. (1979). Educational policy legislation [Interview]. *Educational Evaluation and Policy Analysis*, 1 (4), 71-74.

Rein, M. (1983). *From policy to practice.* New York: M. E. Sharp, Inc.

Riesmann, C. K. (1993). *Narrative analysis.* Newbury Park, CA.: Sage Publications.

Sacken, D. M. & Medinaa, M. (1990). Investigating the context of state-level policy formation: A case study of Arizona's bilingual education legislation. *Educational Evaluation and Policy Analysis*, 12(4), 389- 402.

Sadler, D.R. (1980). Conveying the findings of evaluative inquiry. *Educational Evaluation and Policy Analysis*, 2 (1), 53-57.

Sadler, D. R. (1985). Evaluation, policy analysis and multiple case studies: Aspects of focus and sampling. *Educational Evaluation and Policy Analysis*, 7(2), 143-149.

St. Pieere, R. G. (1982). Follow through: A case study in meta evaluation research. *Educational Evaluation and Policy Analysis*, 4(1), 47-55.

Schubert, W. H. (1980). Recalibrating educational research: Toward a focus on practice. *Educational Researcher*, 9 (1), 17-24, 31.

Select Committee on Children, Youth and Families. (1985, August). *Opportunities for Success: Cost Effective Programs for Children.* (Washington, D.C.: House of Representatives.) (H2 - 385, House Annex #2, 20515)

Shapiro, J. Z. (1985). Evaluation research and program evaluation : Where are we and where do we need to go. *Educational Evaluation and Policy Analysis*, 7(3), 245-248.

Shapiro, J. P., Secor, C., & Butchart, A. (1983). Illuminative evaluation: Assessment of the transportability of a management training program for women in higher education. *Educational Evaluation and Policy Analysis*, 5(4), 465-471.

Shermerhorn, G.R. & Williams, R.G. (1979). An empirical comparison of responsive and preordinate approaches to program evaluation. *Educational Evaluation and Policy Analysis*, 1 (3), 55-60.

Slavin, R. E., & Madden, N. A. (1991). Upping the stakes: Program improvement guidelines to reward appropriate practice. *Educational Evaluation and Policy Analysis*, 13, (4), 369- 379.

Smith, N, L. (1980). Federal research on evaluation methods in health, criminal justice, and the military. *Educational Evaluation and Policy Analysis*, 2(4), 53- 59.

Smith, N. L. (1992). [Review of Evaluation and education: At quarter century]. *Educational Evaluation and Policy Analysis*, 14, (1), 93-97.

Smith, N.L. (1982). The context of evaluation practice in state departments of education. *Educational Evaluation and Policy Analysis*, 4 (3), 373-386.

Smith, L. M., Dwyer, D. C., & Prunty, J. J. (1981). Observer role and field study knowledge : An essay review of usable knowledge and SAFARI I. *Educational Evaluation and Policy Analysis*, 3(3), 83-90.

Stake, R. E. (1979). Counterpoint: More subjective! *Educational Evaluation and Policy Analysis*, 1(1), 46-47.

Stake, R. E. (1981a). Persuasions, not models. *Educational Evaluation and Policy Analysis*, 3(1), 83-84.

Stevenson, H. W. & Siegel, A. E. (Eds.). (1984). *Child development and social policy*. Chicago, IL.: University of Chicago Press.

Stonehill, R. M., & Groves, C. L. (1983). U.S. Department of Education policies and ESEA Title 1 evaluation utility: Changes in attitudes, changes in platitudes. *Educational Evaluation and Policy Analysis*, 5(1), 65-73.

Stufflebeam, D. L., & Webster, W. J. (1980). An analysis of alternative approaches to evaluation. *Educational Evaluation and Policy Analysis*, 2(3), 5-20.

Swan, W. W. (1984). State policy and service achievements in early childhood special education. *Educational Evaluation and Policy Analysis*, 6(4), 425-434.

Szanton, E. S. (1995). "Once upon a time..." public policy and families with infants and toddlers. *Insights from the center for infants and parents*, 2 (1), 7-10 [Teachers College, Columbia University, New York]

Takanishi. R. (1977). Federal involvement in early education (1933-1973): The need for historical perspectives. In L. G. Katz (Ed.), *Current Topics in Early Education: Volume 1* (pp.139-163). (Norwood, NJ: Ablex Pub. Corp.)

Tannen, D. (1984). *Conversational style*. Norwood, NJ: Ablex.

Tanner, D., & Wallat, C. (1983). Doctor/mother/child communication: Linguistic analysis of a pediatric interaction. In S. Fisher & A.L. Todd (Eds.), *The social organization of doctor-patient communication* (pp. 203-220). Washington, DC: Center for Applied Linguistics.

Tannen, D., & Wallat, C. (1993). Interactive frames and knowledge schemas in interaction. In D. Tannen (Ed.), *Framing in discourse* (pp. 57-76). New York: Oxford University Press.

Thompson, B. (1981). Book review: Utilization of evaluation information. *Educational Evaluation and Policy Analysis*, 3(3), 106-107.

Timar, T. (1994). Federal education policy and practice: Building organizational capacity through Chapter 1. *Educational Evaluation and Policy Analysis*, 16, (1). 51-66.

Turnbull, A. P. (1982). School mainstraming: A policy and implementation analysis. *Educational Evaluation and Policy Analysis*, 4(3), 281-291.

vander Ploeg, A. J. (1982). ESEA Title I evaluation: The service of two masters. *Educational Evaluation and Policy Analysis*, 4(4), 521-526.

Wallat, C. (1984). An overview of communicative competence. In C. Rivera (Ed.), *Communicative competence Approaches to Language Proficiency Assessment: Research and application* (pp. 2-33). Avon, England: Multilingual Matters.

Wallat, C. (1991a). Child-adult interaction in home and community : Contributions to understanding literacy. In S.Silvern (Ed.), *Literacy through family, community, and school interaction Vol.5 advances in reading/language research* (pp. 1-36). Greenwich, CT.: JAI Press.

Wallat, C. (1987). Literacy, language and schooling: State policy implications. In D. Bloome (Ed.), *Literacy and schooling* (pp. 291 - 309). Norwood, NJ.: Ablex Publishing Co.

Wallat, C. (1991b, April). *The Concept of "Social Context."* Paper presented at the meeting of the American Educational Research Association, Chicago, IL. (Available as *Educational research directions 1990-1995: Extracts from descriptions of primary tasks of OERI Centers from March 1990 RFP and 17 winning proposals.* Washington, D.C.: American Institutes for Research, ERIC Clearinghouse on Tests, Measurement, and Evaluation. [Clearinghouse No. TMO 17371])

Wallat, C. (1989). *Policymaking and linguistic analysis in education.* Unpublished manuscript. Florida State University Department of Educational Foundations and Policy Studies.

Wallat, C., & Green, J. (1982). Construction of social norms by teachers and children: The first year of school. *The social life of children in a changing society* (pp. 97-122). Hillsdale, NJ: Lawrence Erlbaum Associates.

Wallat, C., & Piazza, C. (1991). Perspectives on production of written educational policy reports. *Journal of Education Policy*, 6, (1), 63-84.

Wallat, C., & Piazza, C. (1989). Policy implications related to teaching and learning mathematics and science. In C. Emihovich (Ed.), *Locating learning: Ethnographic perspectives on classroom research* (pp. 240 - 262). Norwood, NJ: Ablex Pub. Corp.

Wallat, C., & Piazza, C. (1988). The classroom and beyond: Issues in the analysis of multiple studies of communicative competence. In J. L. Green & J O. Harker (Eds.), *Multiple perspective analysis of classroom discourse* (pp. 309 - 342). Norwood, NJ: Ablex Pub. Corp.

Wallat, C., & Steele C. I. (1995a). *Dilemmas of family participation in comprehensive services.* Manuscript submitted for publication.

Wallat, C., & Steele, C. I. (1995b). *The call to pay attention to family diversity: Constructing a response.* Manuscript submitted for publication.

Warfield, M. E. (1994). A cost effectiveness analysis of early intervention services in Massachusetts: Implications for policy. *Educational Evaluation and Policy Analysis*, 16, 1, 87-99.

Wehmeyer, L. B. (1979). Program evaluation for policy- setting. *Educational Evaluation and Policy Analysis* 1(5), 66-71.

Weiss, C. H. (1980). [An EEPA Interview with Carol H. Weiss]. *Educational Evaluation and Policy Analysis*, 2 (5), 75-79.

Wells, F. G. & Peterson, K. D. (1992). External pressures for reform and strategy formation at the district level: Superintendents' interpretations of state demands. *Educational Evaluation and Policy Analysis*, 14 (3), 241-260.

Wildavsky, A. (1977). Principles for a graduate school of public policy. *Urban Analysis*, 4, 3-28.

Williams, D. D. (1986). Naturalistic evaluation: Potential conflicts between evaluation standards and criteria for conducting naturalistic inquiry. *Educational Evaluation and Policy Analysis*, 8(1), 87- 99.

Winfield, L. W. (1991). Lessons from the field: Case studies of evolving schoolwide projects. *Educational Evaluation and Policy Analysis*, 13 (4), 353-362.

Wirt, F. M., & KIRST, M.W. (1972). *The political web of american schools.* Boston: Little, Brown & Co.

Wirt, F., Mitchell D., Marshall, C. (1988). Culture and educational policy: Analyzing values in state policy systems. *Educational Evaluation and Policy Analysis*, 10 (4), 271-284.

Wolf, R. L. (1979). The use of judicial evaluation methods in the formulation of educational policy. *Educational Evaluation and Policy Analysis*, 1(3), 19-37.

Wood, K. C., Peterson, S. E., Degracie, J. S., & Zaharis, J. K. (1986). The jury is in: Use of a modified legal model for school program evaluation. *Educational Evaluation and Policy Analysis*, 8(3), 309- 315.

Wortman, P. M. (1982). Book review : Evaluating with validity. *Educational Evaluation and Policy Analysis*, 4(1), 122-125.

16

About the Authors

Cynthia Wallat

Department of Educational Foundations and Policy Studies
306 Stone Building
Tallahassee, FL. 32306 4070

E-Mail: cwwallat@garnet.acns.fsu.edu

Cynthia Wallat is Professor of Social Sciences and Education in the Department of Educational Foundations and Policy Studies at Florida State University. Her research emphasis and publications address: socialization and language; comparative social policy; and institutional and professional development. Her interest in relating language and policy centers on demonstrating how attention to forms of communication and community can address the known and unknown about diversity in and out of school.

Carolyn Piazza

Carolyn Piazza is Associate Professor of Reading / Language Studies in the Department of Educational Theory and Practice. Her teaching and research emphasis includes: written discourse forms and functions; writing as a social process; writing development; and assessment. Her interest in relating language research to policy work includes analysis of the production of written education policy reports and language and equity issues, including women's styles of writing.

Copyright 1997 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://olam.ed.asu.edu/epaa

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey@olam.ed.asu.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin
John Covaleskie
Northern Michigan University
Alan Davis
University of Colorado, Denver
Mark E. Fetler
California Commission on Teacher Credentialing
Thomas F. Green
Syracuse University
Arlen Gullickson
Western Michigan University
Aimee Howley
Marshall University
William Hunter
University of Calgary
Daniel Kallós
Umeå University
Thomas Mauhs-Pugh
Rocky Mountain College
William McInerney
Purdue University
Les McLean
University of Toronto
Anne L. Pemberton
apembert@pen.k12.va.us
Richard C. Richardson
Arizona State University
Dennis Sayers
University of California at Davis
Michael Scriven
scriven@aol.com
Robert Stonehill
U.S. Department of Education

Greg Camilli
Rutgers University
Andrew Coulson
a_coulson@msn.com
Sherman Dorn
University of South Florida
Richard Garlikov
hmwkhelp@scott.net
Alison I. Griffith
York University
Ernest R. House
University of Colorado
Craig B. Howley
Appalachia Educational Laboratory
Richard M. Jaeger
University of North Carolina--Greensboro
Benjamin Levin
University of Manitoba
Dewayne Matthews
Western Interstate Commission for Higher Education
Mary P. McKeown
Arizona Board of Regents
Susan Bobbitt Nolen
University of Washington
Hugh G. Petrie
SUNY Buffalo
Anthony G. Rud Jr.
Purdue University
Jay D. Scribner
University of Texas at Austin
Robert E. Stake
University of Illinois--UC
Robert T. Stout
Arizona State University

archives | abstracts | editors | board | submit | comment | subscribe | search

EPAA • enter the archives • browse the abstracts • the editors • the edit board
• submit article • submit commentary • search • subscribe
volume: • 1 • 2 • 3 • 4 • 5 • 6 • 7

This article has been retrieved 1348 times since July 29, 1997.

Education Policy Analysis Archives

Volume 5 Number 16 July 29, 1997 ISSN 1068-2341

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass Glass@ASU.EDU

College of Education

Arizona State University,Tempe AZ 85287-2411

Copyright 1997, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

Staffing Up and Dropping Out: Unintended Consequences of High Demand for Teachers

Mark Fetler

Abstract

Growing public school enrollment and the need to maintain or improve service to students has increased the demand for teachers, perhaps more rapidly than existing sources can accommodate. While some schools recruit well qualified teachers by offering higher salaries or better working conditions, others may satisfy their need for staff by relaxing hiring standards or assigning novice teachers to difficult classrooms. Schools' hiring policies have consequences for student success. Dropout rates tend to be higher where faculties include a greater percentage of minimally educated teachers or teachers with little experience. The relationship between dropout rate and teacher qualifications is independent of student poverty, school size, and location. A proposed strategy to reduce dropout rates is to encourage higher preparation and employment standards, and to provide appropriate classroom assignments, mentoring, and support for new teachers.

A systemic view of public schools, while looking for stable recurring processes, also recognizes the law of unintended consequences. For example, public school enrollment growth has stimulated the need to prepare and hire more teachers. Given compulsory school attendance laws, parent expectations, desired student-teacher ratios, and contractual limits on work loads, schools must hire enough teachers to keep pace with growing enrollment. Schools may respond to increased need either by offering incentives or by relaxing standards. Of course, schools that successfully entice more desirable candidates with attractive salaries or good working conditions make hiring more difficult for others. Similarly, state agencies face pressure to ease the standards and regulations for teacher preparation and credentialing when confronted with lobbying from school boards, administrator groups, and teacher organizations. State agencies, schools, colleges, and universities interact, in a loosely defined supply system which operates to staff schools. One goal of the system is to maintain traditionally expected or legally required student-teacher ratios. Although the quality of preparation programs is an important aspect of the supply system, meeting school demands for an adequate number of classroom teachers, arguably is a primary goal.

The price of preparation tends to constrain the number and quality of teachers produced by the supply system. Significant increases in the number of teachers prepared entail greater outlays for training, facilities, and administration, particularly in public universities and schools where the bulk of teacher training takes place. The additional coursework and field experience associated with more rigorous and thorough teacher preparation generally requires increased costs. Understandably, there is resistance from prospective teachers and policymakers to paying more for teacher preparation if it requires spending less on competing priorities.

Educators have long debated systems for instruction in relation to indicators of instructional context, processes, and outcomes. (Levin, 1974; Murnane, 1987; Office of Educational Research and Improvement, 1988; Shavelson, McDonnell and Oakes, 1989; and Porter, 1991) These systems sometimes underlie proposals for structural reforms of public education which aim at particular goals. Conventional goals for student success include academic achievement, citizenship, or job preparation. Persistence in school to a diploma is a minimum goal for at-risk students. The teacher is an essential component of these instructional systems. If it is true that more skilled teachers contribute more effectively to student success, then the education and experience of teachers available for hire is an important aspect of instructional systems.

While relaxing employment standards at the state or local level in response to high demand addresses reduces tension caused by shortages, it raises the likelihood of hiring less well qualified teachers, with possibly less desirable consequences for student success. This paper examines teacher education and experience in relation to student dropout rates. It reviews the relationship between student enrollment and teacher demand. It briefly surveys existing research on the influence of teacher experience and education on student achievement, and on factors influencing student decisions to drop out of school. Actual California high school dropout rates are analyzed and discussed in relation to measures of school size, location, growth, student poverty, teacher education, and experience.

Teacher Supply and Demand

Burgeoning student enrollment is an issue for public schools, influencing not only the need for adequate facilities, but perhaps even more importantly the need for well-qualified teachers. Nationally, NCES (1996) estimates that total K-12 enrollment will grow about 10 percent from 49.8 million in 1994 to 54.6 million by 2006. In California public school enrollment will rise over 18 percent from 5.4 million students in 1995 to 6.4 million ten years later. (See Note 1 in the Appendix.) California's enrollment growth, coupled with statewide efforts to reduce class sizes, is spurring the demand for teachers. An estimated 300,000 teachers will be working in 2005, compared to 232,000 actually employed in 1996, an expansion of 29 percent.

One facet of teacher supply and demand relates to the skills and abilities expected of teachers. (Darling-Hammond and Hudson,1990; Reynolds,1991; National Commission on Teaching and America's Future,1996; Ashton,1996; Education Week, 1997) The institutional and personal resources needed to develop those skills influence the rate at which teachers can be prepared. While teaching expertise is a goal of preparation, usually a credential requires an academic degree and coursework. Although satisfying the requirements may not guarantee competent performance, it is intended to provide assurance that a teacher is prepared for the classroom. Virtually all public school teachers in the United States have at least a bachelor's degree, and a majority possess an advanced degree. (NCES 1995b) The trend is toward higher levels of education. In 1971 28 percent of public school teachers possessed a master's, specialist, or doctoral degree. Twenty years later 53 percent of teachers had an advanced degree.

California's degree and coursework requirements for a preliminary teaching credential generally resemble those of many other states. (National Association of State Directors of Teacher Education and Certification, 1996). Unlike a number of states, California requires a Bachelor's degree in a subject other than professional education. Additionally, California teachers complete a one year preparation program, which provides training in educational principles and teaching strategies. Those who seek a clear credential must fulfill additional course requirements and a fifth year of educationally related study after the Bachelor's. California's requirements allow several routes to a credential. Some candidates complete the Bachelor's degree first, then complete the preparation program as a graduate student. Others work the preparation program and course requirements into their Bachelor's degree in order to receive a preliminary credential. These teachers complete the fifth year of study and remaining requirements within the next five years. Most teachers who transfer from outside of California receive a temporary credential based on completion of a Bachelor's degree and a professional preparation program. Career changers with at least a Bachelor's degree and competence in their subject of instruction may work as paid teaching interns while they receive support and training in pedagogy from school districts or universities.

A second facet of an analysis of teacher supply and demand is the flow of people into and out of public school employment. Sources of credentialed teachers include college and university preparation programs and re entrants from the reserve pool of previously employed teachers. Other sources are school district and university programs to facilitate the mid career transition of people into teaching from jobs in other industries or the military. Nationally, schools are filling an increasing proportion of vacancies with inexperienced applicants. (NCES 1995b) From 1988 to 1991 public schools hired more first-time teachers and fewer reentrants or transfers. Reentrants comprised 33 percent of hires in 1988, compared to 24 percent in 1991. First-time teachers made up 31 percent of hires in 1988, compared to 42 percent in 1991. Teachers who transfer from other schools or return to a school have more experience, but receive higher salaries than first-time teachers. First-time teachers earn less, but are more likely to leave the profession.

Drains on the pool of employed teachers include retirement and migration into other occupations. Nationwide, between 1990-91 and 1991-92 about 5 percent of teachers left teaching, including retirees. (NCES 1995a) Teachers with less full-time teaching experience were more likely to leave. Some 17 percent of those with less than one year of full-time left teaching, compared to 8 percent of those with one year of experience, 7 percent of those with two years of experience, and 6 percent of those with three years of experience.

Some schools have more turnover than others. (NCES 1995c) Smaller schools experience higher teacher attrition. Based on 1990-91 data, schools with less than 300 enrollment had 10.3 percent turnover, compared to 8.2 percent for those with 300-599 enrollment, and 7.7 percent for those with over 600 enrolled students. Lower salaries and benefits may be a factor in this relationship. Small schools offer teachers less compensation than larger schools. For example, small public schools paid teachers an average salary of $35,317, compared to $42,421 paid by large schools. Student poverty is associated with teacher turnover. Schools with over 50 percent of students receiving free or reduced price lunches had teacher turnover rate of about 10 percent, compared to an 8 percent rate for schools with lower proportions of such students.

Interaction of Preparation and Flow

The relationship between teacher preparation and flow is complex. Credential requirements restrict access to the teaching profession. Other conditions remaining equal, higher standards reduce the number of teachers available for work. One way to meet increased demand is to relax the requirements, reducing the time and cost required to become a teacher. For example, when there are too few credentialed applicants, California school districts use emergency permits to hire individuals who lack some requirements for a credential, usually proof of competence in their subject(s) of instruction or pedagogy. (Hart and Burr, 1996) In recent years emergency permits have become more popular. A risk of this increased popularity is that less well prepared teachers may be less effective in their jobs or more prone to attrition.

States have sought to increase the supply of teachers by setting up alternatives to traditional training programs. Zumwalt (1996) describes alternative certification as easing entry requirements, minimizing preparation needed prior to paid teaching, and emphasizing on-the-job training. Proponents portray these programs as attracting higher-ability, more diverse, experienced people with subject matter majors. (Ashton, 1991; Dill 1996; Feistritzer, 1994; Haberman, 1992) Zumwalt cautions that it is difficult to generalize about the success of alternative programs. The recruitment, preparation and retention of teachers is complex. The underlying assumptions are debatable th at the knowledge base of teaching is minimal, that schools can supply the needed mentoring, and that teaching is a craft best learned on the job. Alternative approaches assume that school staffs, already criticized for not meeting the needs of students, have the time, energy, and resources, to support unprepared novice teachers. Plausibly, the success of alternative approaches depends on the extent to which novice teachers actually receive needed support and obtain classroom assignments appropriate to their abilities.

Student Performance

Teacher skills and ability influence student achievement. Greenwald, Hedges and Lane (1996) reviewed a number of studies of the relationship between school inputs and student outcomes. School resource variables which described teacher ability, teacher education, and teacher experience were strongly related to student achievement. On the other hand, Hanushek's (1996) synthesis of research studies found mixed support for a relationship between school resources and achievement. Although Hanushek did not detect a clear pattern, measures of teacher experience were more consistently related to achievement than measures of teacher education. Ashton (1996) notes that teachers with regular state certification receive higher supervisor ratings and student achievement than teachers who do not meet standards. Teachers without preparation have trouble anticipating and overcoming barriers to student learning, and are likely to hold low expectations for low-income children. Ashton suggests that states which reduce certification requirements or permit the hiring of teachers who do not meet certification standards, worsen inequities in the quality of education offered to low income children. For example, some alternative certification programs have minimal training requirements with most teachers placed in economically disadvantaged urban schools.

Student decisions to drop out, thereby delaying or precluding a high school diploma, represent a facet of performance distinguishable from achievement. While high school graduation is a student attainment, it is not a measure of learning. Similarly, while dropping out represents a kind of failure, neither is it a direct measure of achievement. Educators have speculated on the relationship between instructional systems and student dropout rates. During the 1980s California and other states implemented broad curricular and structural reforms aimed at more rigorous academic standards. McDill, Natriello, and Pallas (1985) and Hamilton (1986), among others, speculated that higher standards might result in higher achievement for some students, at the cost of a narrower curriculum and increased chances of dropping out for at-risk students. Although individuals make decisions to leave school in response to particular circumstances (Ekstrom, Goertz, Pollack, and Rock, 1986; Rumberger, 1987; Venezky, Kaestle, and Sum, 1987), they do so under the influence of a school environment. For example, school size and poverty are correlated with dropout rates. (Cibulka, 1986; Toles, Schulz, and Rice, 1986; Pittman and Haughwout, 1987; Fetler, 1989) Overall, the national dropout rate is declining, but is higher in large urban districts. (Schwartz,1995; Coley,1995)

In California Guthrie and Kirst (1988) found that school reforms resulted in a narrowing of the curriculum without an increased risk of dropping out. Between 1981 and 1986 there were statewide increases in academic enrollments, balanced by declines in remedial courses and electives. Schools that successfully implemented the reforms tended to focus on an improved learning environment, heightened concern for all students, teacher collegiality, and teacher and site administrator participation in designing reform implementation activities. Most of these schools took steps to help at-risk students, and did not experience increased dropout rates.

There is large variation in educational attainment in California's adult population. According to the 1990 U.S. Census, 25 percent of Californians 18 or older did not graduate from high school, and 23 percent have just a high school degree. (California Postsecondary Education Commission, 1996) This figure is likely related to California's high rates of immigration compared to many other states, as well as school dropout rates. Annual dropout rates for grades 9 through 12 declined in California from 5.2 percent in 1991 to 4.4 percent in 1994. A comparable national dropout rate was 5.3 percent in 1994.

Method

The unit of analysis was the school. The three types of information analyzed related to schools, teachers, and students. The data were collected from mandated annual surveys administered by the California Department of Education. (See Note 2 in the Appendix for details.) Even in large high schools, employment needs and actions can vary considerably from year to year. In order to obtain relatively stable estimates of teacher education and employment, four year averages of the study measures were computed using data collected from 1993 through 1996. All analyses were weighted by the average number of teachers employed in the school in order to accommodate variation in the size. California had 805 regular high schools serving an average 1.3 million students per year. The majority (N = 749) of these schools offered instruction in grades 9 through 12, although various other grade configurations were represented, most commonly 10-12, or 7-12.

Approximately 600 alternative high schools serving about 100,000 students per year were excluded from the study. Typically, these alternative schools have small enrollments and do not offer the academic curriculum needed to attend California's public universities. Reasons for referral to an alternative school could include an unstable home environment, emotional difficulties, pregnancy, etc. Alternative schools diverge from regular schools in serving a population of students with different needs and providing different kinds of services.

Measured school characteristics included enrollment and location. The average student enrollment was an indicator of school size. Federally derived categories of school location provided the basis for categorizing school location as Large City, Medium City, Urban Fringe, or Rural. The research literature links both of these measures with school dropout rates, and they are included primarily as controls for the teacher education and employment variables.

The school level measures of teacher characteristics included annual growth in the number of employed teachers, the percent of new first-time teachers, the percent of teachers with only a Bachelor's degree, the average number of years of education, and the estimated number of years of experience of the teaching staff. Schools with increasing or declining student enrollment adjust the number of teachers they employ in order to meet the actual need for instruction. High growth schools experience relatively high demand for teachers both through the need to replace teachers who might ordinarily leave or retire, and the need to augment their teaching staff to accommodate added students. The average net annual growth in the size of the teaching staff is an indicator of the stability of the faculty. The percent of new first-time teachers and the average years of experience are indicators of teacher experience. The percent of teachers with only a Bachelor's degree and the estimated number of years of education are indicators of teacher educational background.

The two measures of student characteristics are the percent of students covered by the federal Aid to Families with Dependent Children (AFDC) program, and the annual dropout rate. AFDC is the percentage of students in the school's attendance area who are enrolled in either public or private schools and who are from families receiving aid. As an indicator of poverty AFDC often correlates with student achievement (White, 1982), and functions in this study as a control variable. The annual dropout rate is an indicator of student performance. The annual dropout rate estimates the percent of students who leave during the course of a year, and is smaller than a cumulative rate which estimates the percent of a cohort leaving over a period of years. Student achievement uniformly measured by an objective test is another possible measure of student performance, but is not currently available in California. One other performance measure was investigated, the percent of students completing the academic course sequence needed to attend a public university. Unfortunately the coursework measure lacked sufficient reliability to warrant further analysis. One reason for the unreliability may be that the course content associated with a specific title can vary from school to school.

The descriptive and correlational techniques used in this study permit informed speculation about relationships among the phenomena measured by the study variables. Of course, these techniques by themselves do not justify conclusions regarding cause and effect. Although the data describe a span of four years, the analyses are cross-sectional, and do not permit the examination of change over time, which is often needed to support causal inference.

Results

Table 1 illustrates the differences among schools in various locations. Perhaps the most striking result is the contrast between schools located in large cities and those in rural areas. Rural school dropout rates are less than a third of large city schools, while student enrollment, the percent of teachers with only a Bachelor's degree, and the percent of students in poverty is less than a half of large city schools. Compared to the others, large city schools have higher dropout rates, larger student enrollments, a higher percent of students in AFDC families, and a larger percentage of teachers with only a Bachelor's degree. Rural schools differed from the others, although not so consistently or greatly as the large city schools. Rural schools had teachers with fewer years of experience, and smaller student enrollments. Rural schools also had teachers with fewer years of education than the medium city and urban fringe schools. The findings with regard to teacher experience and education in rural schools could produce an expectation that rural schools would also have a larger percentage of teachers with only a Bachelor's degree. Contrary to this expectation, teachers who possess only a Bachelor's degree make up a smaller percentage of the faculty in rural locations than in other areas.

Table 1

Profile of Schools by Location

Measure	Large City	Medium City	Urban Fringe	Rural
Number of Schools	145	106	418	136
Dropout Rate	7.6(b)	3.3(a)	2.5(a)	1.9(a)
Enrollment	2,347(b)	2,025(a,b)	1,945(a,b)	912(a)
Faculty Growth	1.4	0.6	0.9	0.9
Percent New Teachers	5.7	5.0	5.0(a)	5.4
Years of Education	5.5	5.6(a,b)	5.7(a,b)	5.4
Percent B.A.	18.6(b)	9.4(a)	9.4(a)	8.0(a)
Years of Experience	16.2(b)	16.5(b)	16.5(b)	15.0(a)
AFDC	23.2(b)	15.6(a)	12.4(a)	11.5(a)

Note: Values denoted by (a) differ from those of large city schools, and values denoted by (b) differ from those of rural schools, ($p < .05$), using the Tukey-Kramer HSD comparison method.

Table 2 displays statewide means, medians, and standard deviations. Evidence that three measures are skewed is the difference between the mean and median of the percent of faculty with only a Bachelor's, the percent AFDC, and the dropout rate. These differences indicate that the distributions of schools are skewed so that more schools have lower values of these measures than higher values. The lopsided shape of these distributions probably contributes to lowering the related correlations and regression coefficients obtained in the following analyses. The positive mean and median of the faculty growth indicator is consistent with the overall growth in California's enrollments. Even so, about one third of high schools reduced the size of their faculty over the four years covered by this study.

Table 2

Means, medians, and standard deviations of student, school and teacher measures.

Measure	Median	Mean	Standard Deviation
Student Enrollment (ENR)	2,112	1,983	853
Urban Location (URB)	0	0.23	3.4
Faculty Growth (FGR)	0.6	1.0	3.2
Percent New Teachers (PNT)	4.7	5.2	3.2
Years of Education (YED)	5.6	5.6	0.3
Percent Bachelor's (PBA)	6.6	11.4	11.5
Years of Experience (YEX)	16.1	16.3	2.9
Percent AFDC	11.4	15.3	12.3
Dropout Rate (DOR)	2.1	3.8	3.7

Table 3 shows that school average dropout rates are moderately correlated with all study variables, except the annual growth of the number of faculty. A variable reflecting the urban location of a school was coded with a value of one if the school was situated in an urban setting and coded zero otherwise. Traditionally, school size and poverty are correlated with dropout rates, which is replicated in this study. The average number of years of teacher education and experience are negatively correlated with the dropout rate, so that schools with more highly educated and experienced teachers tend to have fewer dropouts. The percent of teachers with only a Bachelor's degree and the percent of new teachers are positively correlated with the dropout rate, suggesting that schools with minimally educated, novice teachers tend to have more dropouts. Consistent with expectation, years of teacher experience and education are positively associated with one another, and negatively associated with the percent of new teachers and percent of teachers with only a Bachelor's. That is, schools with more highly educated and experienced teachers tend to have fewer novice and minimally educated faculty.

Table 3

Correlations of student, school and teacher measures.

	ENR	FGR	PNT	YED	PBA	YEX	AFDC	DOR
URB	.25	.07	.10	-.21	.34	-.03	.34	.51
ENR		.16	.02	.04	.21	-.10	.26	.39
FGR			.26	.08	.13	-.18	-.09	.10
PNT				-.23	.33	-.43	.09	.26
YED					-.57	.36	-.14	-.25
PBA						-.21	.19	.44
YEX							-.21	-.20
AFDC								.51

Note: Correlations with an absolute value greater than .10 are statistically significant, ($p < .01$).

Even though the indicators of teacher education and experience are significantly correlated with the dropout rate, it is conceivable that school size, location, and AFDC account for both the teacher characteristics and the dropout rate. For example, larger schools appear both to hire more teachers with only a Bachelor's degree and to have higher dropout rates. Perhaps school size mediates the relationship between teacher education level and the dropout rate. AFDC is correlated with the teacher education and experience variables, except for percent of new teachers, and could mediate those relationships. If poverty and school size can explain teacher education and experience as well as dropout rates, it may be that teacher characteristics have little effect of their own.

The multiple regression analysis in Table 4 helps to assess the influence of each indicator on the dropout rate independently of the influence of other measures. A stepwise regression analysis identified four variables which contributed significantly to an explanation of dropout rate. In order of entry these variables were: AFDC, urban location, the percent of teachers with only a Bachelor's, school enrollment, and the percent of new teachers. R-square for this analysis was .50, which represents the proportion of variance in school dropout rates that can be accounted for by the four predictors. While this value of R-square is statistically significant, it is possible that greater values could be obtained by using additional or more precise information about students, teachers, or schools in the analysis. The standardized betas permit a comparison of the importance of these variables in predicting dropouts. AFDC appears to have the greatest impact, with a change of one standard deviation in AFDC related to a change of 2.6 standard deviations in the predicted dropout rate. Urban location of the school trailed AFDC as a predictor of dropout rates. Percent of teachers with only a Bachelor's degree and enrollment were about equally important, followed by the percent of new teachers.

Table 4

Regression Analysis of Dropout Rate

Measure	Parameter Estimate	Standardized Beta
Intercept	-2.21*	0
AFDC	0.11*	2.6
Urban Location	2.64*	2.2
Percent B.A.	0.07*	1.6
Enrollment	0.001*	1.6
Percent New Teachers	0.19*	1.1

*Significant (p < .001). R-square = 0.50

Table 5 displays a profile of two groups of schools identified as in either the top or bottom ten percent with regard to dropout rates. There were 80 schools in each group. The dropout rate of schools in the low group was about one-fortieth of those in the high group. The two groups differed markedly in terms of enrollment and AFDC, consistent with earlier results. The profiles show statistically significant differences between the two groups for the percent of teachers with only a Bachelor's and the percent of new teachers. The percent of teachers with only a Bachelor's was about three times larger in the high dropout group compared to the low dropout group.

Table 5

Profile of Schools with High and Low Dropout Rates

Measure	Low Ten Percent	High Ten Percent
Dropout Rate*	0.3	12.6
Enrollment*	1579	2733
Teacher Annual Growth	2.1	2.0
Percent New Teachers*	4.6	7.0
Years of Education*	5.7	5.4
Percent B.A.*	8.5	24.4
Years of Experience	17.1	15.5
AFDC*	6.1	26.9

* Differences between means are significant, (p < .001).

Discussion

This study replicates traditional findings that higher dropout rates are more likely in larger schools and in poor or urban areas. The relationship between poverty and student performance has long been documented. (White, 1982) The correlates of poverty, perhaps including childhood neglect, lack of family or peer support for education, neighborhood crime, etc. work to decrease the chances of success in school. People who are struggling to meet more basic, short term needs of food, shelter, and physical safety probably attend less to academic development, however much it is in their long term interest. Despite cases of exceptionally effective schools and teachers, or determined parents and students, a substantial performance gap persists between schools with disadvantaged and those with more affluent students.

An argument for larger schools is that economies of scale permit more extensive curricula, as well as more efficient use of facilities and equipment than would be possible with smaller enrollments. On the other hand schools are complex institutions and greater size may tend to make schools seem more intimidating, less welcoming, and less supportive to students. Larger organizations may face greater administrative challenges to deliver student services, provide teacher support, insure discipline, and maintain facilities. Large school size may affect more and less advantaged students differently. Those students who are at-risk of failure and dropping out may become discouraged in larger, more impersonal situations, without sufficient guidance and support. High achieving, more advantaged students may be more able to benefit on their own from the offerings of larger schools.

One facet of research on effective schools is to identify factors which help students overcome disadvantages. This study confirms prior findings that teacher experience and education are two such factors. While earlier research emphasized student academic achievement, this study goes further by looking at students who are at-risk of dropping out. Teacher education and experience appear to influence dropout rates. The smaller the proportion of inexperienced teachers who are new to a school, the lower the dropout rate. The smaller the percent of teachers with only a Bachelor's degree, the lower the dropout rate. This influence appears to hold independently of poverty, and school size, and location.

The stepwise procedure did not identify faculty growth, years of teaching, or overall years of education, so that these variables were excluded from the regression analysis. The results of a regression analysis can be sensitive to the inclusion of highly correlated pairs of variables, a situation described as multicollinearity. For example, the percent of teachers with only a Bachelor's and overall years of education are correlated with each other. However, preliminary analyses which included overall years of education in the model did not appreciably change the results for the other variables, suggesting that multicollinearity is not an issue. Conceptually, it appears that these two resource variables, the overall years of education and the percent of teachers with only a Bachelor's, represent distinguishable characteristics of a school faculty.

Notably, years of teaching and years of education are less strongly associated with the dropout rate than the percent of new teachers or the percent with only a Bachelor's degree. One possible reason for the difference may exist in the assignments that some schools typically give new teachers. It is commonly thought th at more experienced teachers with seniority usually obtain more desirable classroom assignments with well behaved, higher achieving students. Novice teachers lacking seniority receive less desirable, more difficult classrooms with lower achieving, at-risk students. Novice teachers are more likely to have minimum levels of education, particularly if the school district has lowered its hiring standards to maintain staffing levels. Arguably, at-risk students are most in need of intensive, skilled instruction, counseling, and support. Less well educated novice teachers are the least capable members of faculty able to provide the services and support needed to forestall student dropout. Ironically, as novice teachers gain experience and education, they also gain the seniority which enables them to opt out of the difficult assignments. While this pattern of assigning classrooms could exist in many schools, the effects are probably more severe in hard to staff schools with a high proportion of disadvantaged, at-risk students. Such unpopular schools probably have difficulty in retaining their more experienced and educated staff who move on to more attractive work sites. With lower levels of education and experience at a school there is less capability to support and train new teachers. The results of the analyses suggest that this assignment pattern and negative consequences may be more prevalent in urban settings.

Conclusion

Schools are complex institutions comprised of distinct yet interdependent systems. These systems can work in alignment to support student success. To the extent the systems are out of alignment they will be less supportive. Of course, some parts of these systems are beyond the control of schools so that perfect alignment is an unrealistic goal. The system which supplies teachers, depends not only on school administration and personnel offices, but also on employee organizations, state regulatory agencies, college and university teacher training programs, and on the labor market decisions of individuals who choose jobs according to their own interests. The instructional system includes not only teachers, curriculum, classrooms, textbooks, and materials, but also the community social and economic context. Parents and peers may support academic values and activities, or undermine them.

Despite the unruly nature of these systems, educators and policy makers at schools, colleges and in government can strive to work together toward a common vision of student success.

> Ideally, new teachers who are hired to maintain or improve student teacher ratios will have the experience and education needed to support the success of all students, including the disadvantaged and those who are at-risk of dropping out. If new teachers lack some these skills, they should receive appropriate classroom assignments, mentoring, and support. Maintaining a student teacher ratio by accepting lower standards for experience and education does not optimally support student success. Assigning difficult classrooms to unprepared, unsupported, novice teachers additionally threatens at-risk students. While accepting lower standards may be expedient in the short term, there are indirect costs including the personal costs to an individual of a decision to drop out and the burden imposed by dropouts on the public. In the long run, students, educators, and the public are better served by insisting on higher standards, and by providing the resources and the will to implement them.

References

Ashton, P. (Ed.) (1991). Alternative approaches to teacher education. Journal of Teacher Education, 42 (2), 82.

Ashton, P. (1996). Improving the preparation of teachers. Educational Researcher. Vol. 25, No. 9, pp. 21-22.

California Department of Education. (1996). Administrative Manual for CBEDS Coordinators and School Principals. Sacramento: Author.

California Postsecondary Education Commission. (1996). Performance Indicators of Higher Education. Sacramento: Author, Commission Report 96-10.

Cibulka, M. (1986). State level policy options for dropout prevention. Metropolitan Education. 2, 30-38.

Coley, R. (1995). Dreams deferred: High school dropouts in the United States. Princeton: Educational Testing Service, Policy Information Center.

Darling-Hammond, L., and Hudson, L. (1990). Precollege science and mathematics teachers: Supply, Demand, and Quality. In Review of Research in Education (Vol. 16). Washington, DC: American Educational Research Association.

Dill, V. (1996). Alternative teacher certification. In J. Sikula (Ed.), Handbook of Research on Teacher Education (2nd edition, pp. 932-957). New York: Macmillan.

Education Week. (1997). Quality Counts: A Report Card on the Condition of Public Education in the 50 States. Author, Vol. XVI, January 22, 1997.

Ekstrom, R., Geortz, M., Pollack, J., and Rock, D. (1986). Who drops out of school and why? Findings from a national study. Teachers College Record. 87 (6), 356-373.

Feistritzer. C. (1994). The evolution of alternative teacher certification. The Educational Forum. 58 (2), 132-138.

Fetler, M. (1989). School dropout rates, academic performance, size and poverty: Correlates of educational reform. Educational Evaluation and Policy Analysis. 11 (2), 109-116.

Fetler, M. (1997). Where have all the teachers gone? Education Policy Analysis Archives. Vol. 5 No. 2. A peer-reviewed scholarly electronic journal available at http://olam.ed.asu.edu/epaa/

Grossman, P. (1990). The making of a teacher. New York: Teachers College.

Guthrie, J. And Kirst M. (1988). Conditions of Education in California. Berkeley, CA: University of California , Policy Analysis for California Education. Policy paper No. PP88-3-2, pp. 125 - 140.

Haberman, M. 1992). Alternative certification: Can the real problems of urban education be resolved by traditional teacher education? Teacher Education and Practice. 8 (1), 13-27.

Hamilton, S. (1986). Raising standards and reducing dropout rates. Teachers College record. 87(3), 410-429.

Hammack, F. (1986). Large school systems/ dropout reports: An analysis of definitions, procedures and findings. Teachers College Record. 87 (3), 324 341.

Hanushek, E. (1996). A more complete picture of school resource policies. Review of Educational Research. Vol. 66, No. 3, pp. 397-409.

Hart, G. And Burr, S. (1996). A State of Emergency ... In a State of Emergency Teachers. Sacramento: California State University Institute for Education Reform.

Levin, H. (1974). A conceptual framework for accountability in education. School Review, 82, (3). pp. 363-392.

McDill, E., Natriello, F., and Pallas, A. (1985). Raising standards and retaining students: The impact of reform and recommendations on potential dropouts. Review of Educational Research. 55, 415-433.

Murnane, R. (1987). Improving education indicators and economic indicators: The same problems? Educational Evaluation and Policy Analysis, 9, (2), pp. 101-116.

National Association of State Directors of Teacher Education and Certification (1996). Manual on Certification and Preparation of Educational Personnel in the United States. Dubuque, Iowa: Kendall Hunt.

National Commission on Teaching and America's Future (1996). What Matters Most: Teaching for America's Future. New York, Teachers College, Columbia University.

National Center for Educational Statistics. (1996). Projections of Education Statistics to 2006. Washington DC: Author.

National Center for Education Statistics. (1995a). Digest of Education Statistics. Washington, D.C.: U. S. Department of Education Office of Educational Research and Improvement.

National Center for Education Statistics. (1995b). America's Teachers Ten Years After "A Nation at Risk." Washington, D.C.: U. S. Department of Education Office of Educational Research and Improvement, NCES 95-766.

National Center for Education Statistics. (1995c). Which Types of Schools Have the Highest Teacher Turnover? Washington, D.C.: U. S. Department of Education Office of Educational Research and Improvement, NCES 95-778.

Office of Educational Research and Improvement. (1988). Creating Responsible and Responsive Accountability Systems: Report of the OERI State Accountability Study Group. Washington, DC.: U. S. Department of Education.

Pittman, R. And Houghwout, P. (1987). Influence of high school size on dropout rate. Educational Evaluation and Policy Analysis. 9 337 - 343.

Porter, A. (1991). Creating a system of school process indicators. Educational Evaluation and Policy Analysis, 13, (1), pp. 13-29.

Rumberger, R. (1987). High school dropouts: A review of issues and research. Review of Educational Research, 57, 101 - 121.

Schwartz, W. (1995). School dropouts: New information about an old problem. Washington, DC: U.S. Department of Education, Office of Research and Improvement. (ED 386515)

Shavelson, R., McDonnell, L., and Oakes, J. (Eds.) (1989). Indicators for Monitoring Mathematics and Science Education: A Sourcebook. Santa Monica, CA: The RAND Corporation.

Toles, T., Schulz, E., and Rice, W. (1986). A study of variation in dropout rates attributable to effects of high schools. Metropolitan Education. 2, 30 - 38.

Venezky, R., Kaestle, C., and sum, A. (1987). The subtle danger: Reflections on the literacy abilities of America's young adults. Princeton, NJ: Educational Testing Service.

White, K. (1982). The relation between socioeconomic status and achievement. Psychological Bulletin, 91, 461-481.

Zumwalt, K. (1996). Simple Answers: Alternative Teacher Certification. Educational Researcher, Vol. 25, No. 8, 40-42.

Appendix

Note 1. Enrollment and Teachers

The actual graded K-12 enrollment and numbers of teachers shown in Table 6 are from the California Basic Educational Data System (CBEDS) of the California Department of Education. Enrollments are published annually in a document entitled "California Public Schools Enrollment." Counts of teachers are published annually in a document entitled "Count of Certificated and Classified Staff in California Public School Districts." The teacher counts reflect certificated staff with classroom assignments, and exclude administrators and pupil services staff.

Table 6

Actual and Projected Enrollment and Teaching Staff

School Year	Graded K-12 Enrollment	Teachers	Class Size Reduction Program
1991-92	5,001,670	219,353	
1992-93	5,089,808	220,871	
1993-94	5,166,261	223,932	
1994-95	5,242,078	228,204	
1995-96	5,367,926	232,488	
1996-97	5,495,075	238,951	259,000
1997-98	5,623,422	244,532	266,000
1998-99	5,737,874	249,509	269,000
1999-00	5,841,535	254,017	274,000
2000-01	5,945,067	258,519	279,000
2001-02	6,052,242	263,179	284,000
2002-03	6,160,231	267,875	289,000
2003-04	6,271,881	272,730	295,000
2004-05	6,392,367	277,969	300,000

The projections of graded public school enrollment are published by the California Department of Finance Demographic Research Unit in a document entitled "K-12 Graded Public School Enrollment by Ethnicity, History, and Projection - 1995 Series." The projections are based on a grade-progression ratio (or cohort survival) projection method and the most recent ten years of historical enrollment data from CBEDS.

The actual average ratio of K-12 pupils to classroom teachers from 1991 92 through 1995-96 is 23 to 1. The projected numbers of teachers from 1996-97 onward assume continuation of the 23 to 1 student teacher ratio. During the fall of 1996 the California Legislature enacted a program, which gives incentives to school districts to reduce class size in three elementary grades. Under this program there is a limit of twenty students in a "class." An estimated 20,000 additional teachers are needed to fully implement this program, which represents about an eight percent increase in the size of the work force. The projected number of teachers under the Class Size Reduction Program is calculated by applying an eight percent increase to the original projections. Over the next ten years, with class size reduction, the teaching workforce should increase in size by 68,000, which is about 30 percent growth. Fetler (1997) provides a more detailed discussion of enrollment growth in relation to the supply of teachers.

Note 2. Sources of Data

The annual Professional Assignment Information Form (PAIF), administered as a part of California Basic Educational Data System, was the source of teacher descriptive measures for each school, including: number of teachers, years of experience, percent new teachers, teacher educational level, percent of teachers with only a Bachelors degree, and growth in teacher staffing levels. School level personnel actions can vary from year to year, depending on local policies, resources, and needs. In order to improve the stability of the employment related measures, four year averages (1993 through 1996) of these measures were computed.

- Years of experience is defined as the total years of public and/or private educational service.
- Percent new teachers is computed as the number of teachers with no previous educational service divided by the total number of teachers in the school.
- Teacher years of education is a computed index. Teachers' responses were coded as follows: Bachelors degree = 4, Bachelors plus 30 semester hours = 5, Masters degree = 6, Masters plus 30 semester hours = 7, and Doctorate = 8. The coded responses were averaged to produce the index.
- The percent of teachers with only a Bachelors degree is computed as the number of teachers with only a Bachelors degree divided by the total number of teachers in the school.
- The growth in the number of teachers was computed as the value of the slope coefficient of a four year linear trend. This coefficient is interpreted as the average yearly change in the number of teachers employed at the school.

CBEDS high school profile data sets were the source of measures of school characteristics, including the school dropout rate and the proportion of graduates meeting the public university course requirements, and the percent of students receiving Aid to Families with Dependent Children. Four year averages (1993 through 1996) of these measures were computed in order to improve stability. These school summary files are available at the web site hosted by the California Department of Education:

http://goldmine.cde.ca.gov/ftpbranch/retdiv/demographics/Demohome.html

The school dropout rate is computed as the number of reported dropouts in grades 9 through 12 divided by the eligible number enrolled. According to CBEDS (1996) a dropout is a person who meets the following criteria:-

- was formerly enrolled in grades 7, 8, 9, 10, 11, or 12-
- has left school for 45 consecutive school days and has not enrolled in another public or private educational institution or school program-
- has not re-enrolled in the school-
- has not received a high school diploma or its equivalent-
- was under twenty-one years of age-
- was formerly enrolled in a school or program leading to a high school diploma or its equivalent.

The percent receiving AFDC is the percentage of students in the school's attendance area who are enrolled in either public or private schools and who are from families receiving Aid to Families with Dependent Children (AFDC). This measure was taken from the high school performance data sets available at the same web address.

The classification of schools into demographic regions was derived from information contained in the 1994-95 Public School Name and Address File disseminated by the National Center for Education Statistics (1996). http://www.ed.gov/NCES/ccd/index.html

This file contains the names and addresses of the 89,151 public schools in the 50 states, District of Columbia, and five outlying areas for 1994-95. Also included on each record is the School's enrollment (membership) and various other codes. These codes comprise school type, lowest and highest grades taught, and school locale.

The locale code is a definition of how the school is situated in a particular location relative to populous areas, based on the school's mailing address. The Code translations are as follows:

1= Large Central City
2= Mid-size Central City
3= Urban Fringe of Large City
4= Urban Fringe of Mid-Size City
5= Large Town
6= Small Town
7= Rural

The definitions for locales are:

- Large City: Central city of a Metropolitan Statistical Area (MSA) with a population greater than or equal to 400,000 or population density greater than or equal to 6,000 people per square mile.
- Mid-size City: Central City of an MSA with a population less than

400,000 and a population density less than 6,000 people per square mile.

- Urban Fringe of Large City: Place within an MSA of a Large Central City and defined as urban by the Census Bureau.
- Urban Fringe of Mid-size City: Place within an MSA of a Mid-size Central City and defined as urban by the Census Bureau.
- Large Town: Town not within an MSA, with a population greater than or equal to 25,000.
- Small Town: Town not within an MSA and with a population less than 25,000 and greater than or equal to 2,500 people.
- Rural: A place with less than 2,500 people and coded rural by the Census Bureau.

Only a small number of high-schools were in the "Small Town" category, which was therefore merged into the "Rural" category. The two "Urban Fringe" categories were a lso combined as were the "Large City" and Large Town" categories after verifying that they were similar in terms of the other study variables.

About the Author

Mark Fetler

California Department of Education
721 Capitol Mall
Sacramento, CA 95814

(916) 657-4267
mfetler@cde.ca.gov

Employment Experience

- Consultant. California Department of Education. 1997 - present. Coordinate the development statewide assessment standards and tests administered under the Golden State Examination Program.
- Consultant. California Commission on Teacher Credentialing. 1995 - 1997. Plan, organize and conduct research on teacher credential examinations and teacher supply and demand.
- Director. Planning, Effectiveness and Accountability Unit, Chancellor's Office, California Community Colleges. 1990 - 1995. Manage accountability task force. Develop federal and state accountability programs.
- Administrator. Educational Planning and Information Center, California Department of Education. 1984 - 1990. Manage and develop K-12 accountability policies and programs.
- Consultant. California Assessment Program, California Department of Education. 1980 - 1984. Conduct research and assessment projects.
- Evaluation Specialist. Northwest Regional Educational Laboratory, Portland, Oregon. 1979 - 1980. Deliver evaluation technical assistance and training to state

education agencies.

- Senior Staff Associate. Western Interstate Commission for Higher Education, Boulder, Colorado, 1978 - 1979. Evaluate research instruments.

Education

B.A., Pyschology, Colorado College, 1972

Ph.D., Psychology, University of Colorado, 1978

Study abroad in Germany at the Universities of Göttingen and Bielefeld

Publications

Authored research in peer-reviewed journals, e.g., *American Educational Research Journal, Educational Evaluation and Policy Analysis, Applied Measurement in Education, Sex Roles*, and the *Journal of Communication*.

Community Service

Volunteer for Elk Grove Unified School District, Board member of the Elk Grove Community Planning Advisory Council, fund raiser for the Strauss Festival of Elk Grove, and member of the Elk Grove Rotary Club.

Copyright 1997 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://olam.ed.asu.edu/epaa

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey@olam.ed.asu.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

John Covaleskie
Northern Michigan University

Andrew Coulson
a_coulson@msn.com

Alan Davis
University of Colorado, Denver

Sherman Dorn
University of South Florida

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
hmwkhelp@scott.net

Thomas F. Green
Syracuse University

Alison I. Griffith
York University

Arlen Gullickson
Western Michigan University

Ernest R. House
University of Colorado

Aimee Howley
Marshall University

Craig B. Howley
Appalachia Educational Laboratory

William Hunter
University of Calgary

Richard M. Jaeger
University of North Carolina--Greensboro

Daniel Kallós
Umeå University

Benjamin Levin
University of Manitoba

Thomas Mauhs-Pugh
Green Mountain College

Dewayne Matthews
Western Interstate Commission for Higher Education

William McInerney
Purdue University

Mary P. McKeown
Arizona Board of Regents

Les McLean
University of Toronto

Susan Bobbitt Nolen
University of Washington

Anne L. Pemberton
apembert@pen.k12.va.us

Hugh G. Petrie
SUNY Buffalo

Richard C. Richardson
Arizona State University

Anthony G. Rud Jr.
Purdue University

Dennis Sayers
University of California at Davis

Jay D. Scribner
University of Texas at Austin

Michael Scriven
scriven@aol.com

Robert E. Stake
University of Illinois--UC

Robert Stonehill
U.S. Department of Education

Robert T. Stout
Arizona State University

archives | abstracts | editors | board | submit | comment | subscribe | search

This article has been retrieved 1438 times since August 25, 1997.

Education Policy Analysis Archives

Volume 5 Number 17 August 25, 1997 ISSN 1068-2341

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass Glass@ASU.EDU.

College of Education

Arizona State University,Tempe AZ 85287-2411

Copyright 1997, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

Academic Freedom, Promotion, Reappointment, Tenure and The Administrative Use of Student Evaluation of Faculty: *(Part II)* Views From the Court

Robert E. Haskell
University of New England[1]

Abstract: Though a controversial history of research on the reliability and validity of student evaluation of faculty (SEF) exists, it has not been typically viewed as an infringement on academic freedom, promotion, reappointment, and tenure rights. As a consequence, legal aspects of SEF are neither readily apparent, nor available. Unlike academic freedom, tenure, and other issues, which exist as legal categories, SEF as a category is virtually absent in legal compendia on higher education law. The question of its judicial standing is important to any suggestion of abridging faculty rights. In this second of four articles, legal rulings are categorized and abstracted verbatim from cases where SEF is integral to the denial of academic freedom, tenure, promotion and reappointment are reviewed and provide and provide an initial resource of legal ruling on SEF. Seventy-eight findings are summarized from the abstracted textual material.

Table of Contents

443

- **Use of SEF**
- **Legal Rulings on The Reliance and Validity of SEF**
    - ○ Reliance on SEF v. Peer Evaluation
    - ○ Popularity and Effectiveness
    - ○ Transcendent Value of a Professor Over Teaching Quality
- **The Courts' Approach to the General and Psychometric Accuracy/Validity of SEF**
    - ○ Numerical Ranking of Faculty
- **Rulings on The Qualitative and Subjective Assessment of SEF**
    - ○ Acceptance of Administrative Subjective Judgements of SEF Data
    - ○ Use of Qualitative Written Student Comments
    - ○ Mixed Student Comments
- **Variables Not Taken Into Account When Assessing SEF**
    - ○ Student Biases in Affecting SEF
    - ○ Other Variables Not Taken Into Account
- **Courts Impact on Academic Matters**
    - ○ Impact on Academic Standards
    - ○ Methods of Instruction and Academic Freedom
    - ○ Release of SEF To Students and To the Public
- **Conclusion**
- **References**
- **Appendix A: Summary of 78 Findings**
- **Appendix B: Summary of Case Outcomes**
- **Appendix C: Abstracted Case Material**

Unlike the extensive body of conflicting research on the construction, assessment, validation, and interpretation of student evaluation of faculty (SEF), until recently there has been virtually no categorical mention of their impact on academic freedom and tenure rights (Haskell, 1997). When SEF is recognized to have an impact, unlike traditional threats to tenure and academic freedom, it does not appear to warrant serious concern. Moreover, as I noted in my first article, when it is suggested that SEF may impinge on academic freedom, tenure, promotion and reappointment, it is generally considered an attack on either student rights, the process of evaluating faculty performance in general, or both. Unlike academic freedom, tenure, and other educational issues, as a retrieval category SEF is virtually absent in legal compendia on higher education law. Sources citing legal cases on the issues of academic freedom, tenure and promotion do not typically address SEF.

A recent booklet on *"The Law of Teacher Evaluation"* (Zirkel, 1996) contains no mention of SEF cases. Nor does a recent comprehensive legal guide for educational administrators, (Kaplin and Lee, 1995), nor do other reports (Poch, 1993) on the legalities of academic freedom, tenure and promotion. Two other recent sources, however, do address the issue, one very briefly, (Baez, and Centra, 1995), the other somewhat more extensively (Weeks, 1996). This absence is both surprising and important given that courts have recognized for some time "that in recent years increasing weight has been placed upon teaching ability" (Johnson v. University of Pittsburgh, 1977).

In a rare mention of SEF as an infringement on academic freedom, a recent handbook for college administrators, refers to the concept as "a rather novel attack on the use of student evaluations in assessing a faculty member's performance" (Weeks, 1996). It has generally been taken for granted, not only by many administrators but by faculty as well, that SEF is both appropriate and necessary.[2]

There are, however, increasing signs indicating the examination of this issue by faculty.[3] It is important to note at the outset that it is not SEF *per se* that is the issue in this paper, but rather the legal and educational implications of its use in salary, promotion and tenure decisions. Neither is the primary concern with faculty welfare. As I will delineate in a following paper (Part IV), the primary issue is the implications for academic freedom in the service of standards and quality of student instruction. Such signs would suggest a review of legal rulings involving SEF.[4]

In addition to not considering SEF a significant issue in the abridgement of academic freedom, tenure, promotion, and reappointment, a further reason for the paucity of articles on this issue is the difficulty in searching for legal cases. It appears that cases involving student evaluation of faculty are most always only a part of a much wider set of issues regarding denial of academic freedom, promotion, tenure, and reappointment.[5] In addition, it seems that such cases are often settled not in courts of law but by Labor Relation Arbitration Boards rulings. Since rulings on SEF are embedded within other cases, locating them through key word searches is extremely difficult.[6] Accordingly, there exists no systematic cataloging of how courts have specifically viewed SEF in cases involving denial of academic freedom, tenure, promotion, and reappointment.

Because administrative policy regulating the use of SEF has largely evolved in a de facto or default manner, this paper will examine legal reasoning and rulings on SEF. In the evolution of any legal standard, the accumulation of data, judgements, and arguments around an issue need to be coalesced. As the history of legal rights demonstrates, issues not considered to have legal standing only come to have standing after a long and arduous process of analysis and advocacy. It is the purpose of this paper to coalesce the legal rulings on SEF and to thus facilitate this developmental process, and to serve as an initial resource for faculty organizations and other policy makers. In doing so, I would like to make clear that I am not an attorney and approach this paper and court rulings from the legal concept of the "reasonable man" standard.

In the first of what will be four articles on SEF (Haskell, 1997), I suggested that SEF is deceptive regarding its negative implications for higher education. In this second article, based on the content of located cases, I have developed a set of categories and have abstracted the corresponding rulings. More importantly, it should be noted that the issues involved in SEF go beyond problems involved in general faculty evaluation procedures. Delineating the ramifications of SEF provides a concrete vehicle for further understanding the abstract concept of academic freedom.

To render the boundaries of this paper manageable, its focus will be delimited to how courts[7] have addressed SEF issues within various legal challenges to the denial of academic freedom, tenure, promotion, and reappointment by institutions of higher education. It is important to note that there are multiple legal variables that determine an action or final outcome in any given case. For example particular statutory considerations, cause of action, level of prima facie evidence, standards and burden of proof, and level of judicial action, i.e., whether it is pre trial, lower level court, or appellate, etc., (see Kaplin and Lee, 1995, section 1.4.3.6). For my purposes here, I will not be concerned with these variables. Accordingly, this paper will neither be concerned with the validity or wisdom of the legal rulings, nor with the complex reasoning on which the rulings were based. My purpose is to review the general reasoning of the courts on SEF. Only when specifics of a case are contextually warranted in order to clarify the court's view of SEF will contextual specifics be noted. [8] Not only are such cases prima facially complex, but when specific legal definitions (e.g., disparate impact and treatment) and other special Congressional Acts (e.g., EEOC) are superimposed on them, they become logically unwieldy, not just to the non legal scholar, but apparently even to the Courts.[9]

I should note that the collective rulings are an extremely disparate and frequently contradicting corpus. I should note, too, that in presenting this complex legal corpus, in order to avoid any misinterpretation, I felt necessary to quote extensively from the original material rather than paraphrasing and requiring the reader to constantly refer to the Appendix, which is provided as a resource. For ease of reading and convenient referencing, I have placed the quoted legal rulings, along with my own necessary transitional phrases connecting the quotes, within an indented text format Finally, I will be making no distinctions between the various levels of judiciary proceedings.

Use of SEF

Since the 1960's SEF has been increasingly used in decisions on tenure and promotion. In comparing two studies of the same 600 liberal-arts colleges, the author found that the number of institutions using student ratings to evaluate instructors had escalated from 29 per cent to 68 per cent to 86 per cent. The author (Seldin, 1993) noted that no other method of evaluation has approached that level of usage. Another survey found that most business schools now use SEF for decision making, with 95% of the deans at 220 accredited undergraduate schools always making use of them as a source of information (Crumbley, 1995). Two nationwide studies of accounting department Chairpersons, indicated that reliance upon SEF was second only to research publications in professional journals (Yunker and Sterner, 1988). Department chairs and Deans often weigh student ratings heavily in the faculty evaluation process. With the exception of publications, no other method of faculty evaluation has become so sacrosanct.[10] The reason that SEF was initially instituted was for informational feedback to faculty so they might be more aware of student needs.

Not only does SEF have legal standing, but in some states it is mandated for secondary teachers. According to one author, student evaluations in some states have legal weight, and teachers have been fired on the basis of low ratings on their student evaluations. In California SEF is apparently mandatory under law "whenever practicable" (in, Allegre, and Guista, 1993). SEF is used not only in the U.S. but in Australian, Canada, Europe and Great Britain. Unlike in the U.S., however, in other countries SEF by formal questionnaire, despite apparently no formal mandate, is increasingly used, though not weighed as heavily as is information gathered by other means (Husbands and Fosh, 1993). The instrument has not been used just for informational feedback, however. If it was only used for feedback, then SEF would presumably not be a problem.[11] As Cashin (1996), Director of the Kansas State University, Center For Faculty Evaluation and Development, notes, "The higher education rhetoric is almost universal in stating that the primary purpose of faculty evaluation is to help faculty improve their performance. However, an examination of the systems--as used--indicates that the primary purpose is almost always to make personnel decisions. That is, to make decisions for retention, promotion, tenure, and salary increases." Herein lies a nest of problems.

Legal Rulings on The Reliance and Validity of SEF

This section, will examine courts' approach to SEF in the denial of academic freedom, tenure, promotion, and reappointment decisions. In so doing, I have abstracted from located cases the following categories: (1) Reliance on SEF v. Peer Evaluation, (2) Popularity and Effectiveness, (3)Transcendent Value of a Professor Over Teaching Quality, (4) Establishing the General and Statistical Accuracy/Validity of SEF, (5) Numerical Ranking of Faculty, (6) Acceptance of Administrative Subjective Judgements Of SEF Data, (7) Use of Qualitative Written Student Comments, (8)Mixed Student Comments, (9) Student Biases in Affecting SEF, (10) Variables Not Taken Into Account When Assessing SEF, (11)Courts View of Who Sets Academic Standards, (12) Methods of Instruction and Academic Freedom (13) Release of SEF To Students and To The Public.[12]

Within these categories, I have included Canadian court findings.[13] I have chosen to include Canadian cases for three reasons. First, to provide a larger data base illustrating how legal bodies reason and rule on cases involving SEF, second, to provide a wider array of an otherwise relatively small number of U.S. cases; and third, to provide a set of comparative views from a culture and legal system that has both an interesting set of similarities and differences.[14] In two following papers, I will further address and analyze the impact and implications of these legal rulings in relation to (a) psychometric Validity, and (b) academic freedom and instruction.

Reliance on SEF v. Peer Evaluation

One important issue is how the courts view the relative weighting of SEF in administrative decisions of teaching competence. Is it acceptable, for example, to rely heavily on them, or must they be used in conjunction with other evaluative methods?

> In *Johnson v. University of Pittsburgh* (1977), the court noted that it (5) [15] "has placed little reliance on students' surveys....students in a given course rating a teacher, or professor, some of them as excellent, others as terrible and in between, many who say passable, mediocre etc.... we cannot say it was unreasonable for the tenured faculty to consider this along with other matters" (p.1359). (8) "It is also obvious that the court and the administration of universities cannot permit students to exercise a veto over professors who may be world renowned scientists" (p.1366-7). A similar view was expressed in *Yu Chuen Wei v. Vermont State Colleges Faculty Federation* (1995).
>
> In Peters v. Middlebury College (1977), the court gave some weight to an administrative devaluing of a set of positive student evaluations of a faculty that said (2) "The department chair sent a letter to the president of the college, saying, 'The course of action I recommend is not likely to be popular with students who, though they in part recognize her intellectual limitation, are warmly responsive to her enthusiasm, energy, openness and ready human concern' " (p.860).
>
> In Carley v. Arizona Board of Regents, (1987), the court said, (23) "Carley has cited no authority that *relying primarily or solely on student evaluations would be impermissible*. We have found none" (p.1105, italics added).

> In *Guam Federation of Teachers v. The University of Guam* (1990), the Guam Federation of Teachers challenged the use of SEF in tenure and promotion decisions (Blum, 1990). The Board (1) ruled to remove anonymous student evaluations from professors' tenure files, (2) The union said the use of SEF violated the union's contract with the university, (3) which provides that anonymous documents or those "based on hearsay" should not be included in a faculty member's file, (4) The court further ruled that (5) students should be made aware of the purpose and ramifications of their evaluations, and (6) anonymous student evaluations should not be used.
>
> In *Robert Kramer v. The President of the University of British Columbia* (1992), the Board noted that (18) "The most important perceived error in the teaching evaluation, in the opinion of the Board, is the reliance solely upon the student evaluations and written comments...There was no peer review at all; no member of the Department audited any of Dr. Kramer's lectures" (p.10).
>
> In *University of Regina Faculty Association v. University of Regina* (1993) a Canadian Arbitration Board ruled that (3) "With respect to teaching, it is our opinion that the evidence of unsatisfactory performance is very weak indeed... It is important to note that the basis of the comments, particularly the negative ones in the fall of 1992, were written student assessments... [and] Although these assessments are expressly recognized in Art. 17.19 of the collective agreement, to base important career decisions on them only does not seem justified" (p.4). The Board further ruled (4) that tenure decisions could not be based solely on assessments which were completed by students who had never been made aware of the ramifications of their statements. (5) [I]f evaluations are to be used for serious career development purposes those completing them should be aware of the potential consequences of their participation" (p.4) (8) "To base serious career decisions narrowly on student evaluations is not to be encourage... (9) If teaching is to be seriously evaluated for career purposes, whether for positive or negative purposes, it seems incumbent upon Faculties not to rely only on classroom administered evaluations but to broaden the base of assessment" (p.4)
>
> In *Christopher Turner v. The President of the University of British Columbia* (1993), the Board ruled, "(9) while the [Faculty Association] Agreement permits, but does not mandate either student reviews or peer reviews, and the methods of assessment 'may vary', we do conclude that the reliance placed on these very limited student reviews must have been great, since there was no other evaluation referred to. Where there is no other evidence sought, student comments will have an apparent importance and credibility that they may not deserve... (10) We would strongly recommend peer review in the reconsideration which we are requiring" (p.7). The board further noted that (8) "This board has been asked on a number of occasions to pass judgment on the relevance of student evaluations to the [Faculty Association] Agreement criteria for good teaching. Good teaching is an elusive concept. Students may not be good judges during a course; their judgment might be quite different several years later in life" (p.7).

Summary: From these cases, it can be seen that court rulings range from saying that (1) relying primarily or solely on student evaluations is acceptable, to (2) placing little exclusive reliance on SEF, (3) in rare cases SEF can not be permitted to stand in the way of promoting or retaining professors who are excellent in non teaching areas, (4) tenure decisions can not be based solely on SEF by students who have not been made aware of the ramifications of their evaluations, (5) anonymous documents or those "based on hearsay" should not be included in a faculty member's file, (6) students should be made aware of the purpose and ramifications of their evaluations of faculty, (7) anonymous student evaluations should not be used, (8) peer evaluations must also be a part of evaluating teaching.

Popularity and Effectiveness

A significant issue in SEF is the extent to which it measures popularity not teaching effectiveness. Accordingly, it is instructive to see how courts view this issue.[16]

> In *Johnson v. University of Pittsburgh* (1977), the court said, "It is also obvious that the court and the administration of universities cannot permit students to exercise a veto over professors who may be world renowned scientists and yet if the students rate them unfavorably can be terminated at any time because of unpopularity" (p.1366-7).
>
> In *Carley v. Arizona Board of Regents*, (1987), he (8) he maintained his popularity suffered as reflected in his low student evaluations; in *Robert Kramer v. The President of the University of British Columbia* (1992), he maintained that (14) student evaluations were considered from the standpoint of his popularity, not his effectiveness; and in *Brian Maclean v. President of The University of British Columbia* (1991), (35) The Faculty Agreement specified that "Evaluation of teaching shall be based on the effectiveness rather than the popularity of the instructor." Courts have ruled in various directions on this issue.
>
> In *Robert Kramer v. The President of the University of British Columbia* (1992), the board noted (21) "As for the 'popularity vs. effectiveness' debate, a discouraging or hostile attitude is a part of effectiveness as much as it is of popularity" (p.8).
>
> In *Christopher Turner v. The President of the University of British Columbia* (1993), the Board ruled, (8) "while popularity is not competence nor effectiveness, to the extent that it encourages students it has some relation to both" (p.7).

Summary: From these cases, it can be seen that court rulings range from saying that (9) in cases of exceptional research faculty that popularity should not play a role in termination due to teaching, to (10) in normal cases that a measure of popularity is related to teaching effectiveness.

Transcendent Value of a Professor Over Teaching Quality

Despite the importance placed on teaching, there is precedent for both school policy and the courts---under certain conditions---to ignore poor teaching as indicted by SEF.

> In *Johnson v. University of Pittsburgh* (1977), the court said, "It is also obvious that the court and the administration of universities cannot permit students to exercise a veto over professors who may be world renowned scientists" (p.1366-7), noting, "It is obvious that a professor may be possessed of excellent qualifications as a research scientist and not necessarily be able to prove his or her worth as a teacher, concluding that, (9) "in cases where one has an outstanding scientist of national or international reputation, one may decide to promote and give tenure notwithstanding inability to come across as a teacher, this however is not one of those cases" (p.1366-7).
>
> In *Yu Chuen Wei v. Vermont State Colleges Faculty Federation* (1995), (31) Wei's last claim charged that the College violated the Contract by denying her a promotion, even though both her scholarly performance and professional activities were exceptional. Article 22(E) of the College provides for otherwise granting promotion if the President decides that "performance in one of three areas has been exceptional" (p.314). The Board concluded that "Although Grievant had a significant publication record, most of it was developed before coming to Castleton" (p.315). (33) In terms of exceptional scholarship, Dr. Wei maintained she had solved a significant mathematical problem (apparently published). The Board's response was, (34)

> "although Grievant claimed to have solved the Erdos conjecture, [the]Dean reasonably concluded that she had not established that she actually had solved the conjecture. Under these circumstances, and given our consideration of the discrimination issue previously discussed, we conclude that (35) Grievant has not established discrimination. The Colleges reasonably, and based on legitimate reasons, concluded that Grievant had met the tenure standards in this performance area but that her performance was not exceptional" (p.315).
>
> In *Dr. Brian Maclean v. President of The University of British Columbia* (1991), the court said, (34) "while a superior research and publication record cannot overcome a poor teaching record, it might tip the scales where the teaching record was 'on the edge'" (p.10).

Summary: Courts have ruled that (11) the courts and educational administrations can not allow low SEF to stand in the way of promoting or retaining professors who may be world renowned scientists, (12) deemed nationally or internationally exceptional as a researcher, courts have variously ruled that SEF may be disregard and not disregarded, (13) at least in these two cases the courts did not find the faculty exceptional.

The Courts' Approach to the General and Psychometric Accuracy/Validity of SEF[17]

An issue directly related to the relative reliance on SEF for administrative purposes is validity. It would presumably appear from any "reasonable man" standard that the more valid the SEF data in a given case, the more justifiable is the reliance on them.

> In *Johnson v. University of Pittsburgh* (1977), the court said (7) "We have repeatedly approved the use of statistical proof where it reached proportions comparable to those in this case to establish a prima facie case of racial discrimination in jury selection cases . . Statistics are equally competent in proving employment discrimination. We caution only that statistics are not irrefutable. They come in an infinite variety and, like any other kind of evidence they may be rebutted. In short, their usefulness depends on all of the surrounding facts and circumstances" (8) The court further said in Footnote # 20: "Considerations such as small sample size may of course detract from the value of such evidence" (p.1361).
>
> In *Peters v. Middlebury College* (1977), it was maintained that (5) "A professor's value depends upon his creativity, his rapport with students and colleagues, his teaching ability, and numerous other intangible qualities which cannot be measured by objective standards" (p.860).
>
> In *Fields V. Clark University* (1987), the court noted that (10) Fields' "attacks" the university's use of her student evaluations because they were not gathered and evaluated according to accepted standards of scientific polling procedures. In response, the court agreed, saying, "She is probably correct. The use made of the student evaluations in her case, however, followed the practice at the defendant's university in other tenure decisions" (P.671).
>
> In *Cynthia J. Fisher v. Vassar College* (1995), the court noted that (7) "statistical analyses may be a part of a plaintiff's effort to establish discriminatory treatment" (p.1209).
>
> In *Yu Chuen Wei and the Vermont State Colleges Faculty Federation* (1995), the court ruled that (4) "The Court need not consider the accuracy of these administrative determinations, and that (24) tenure criteria "are not drawn with mathematical nicety."

> The board further ruled that (25) "the Dean and the President, both reviewed Grievant's student evaluations carefully. Their failure to take it a step further, and perform a statistical comparison of Grievant's student evaluations with those of other faculty members who have been granted tenure was not arbitrary and was reasonable; (26) Such a comparison is nowhere required by the Contract, [and] (27) we decline to hold such an involved comparison is necessary before a reasonable tenure determination can be made" (p.311).
>
> In *Dr. Brian Maclean v. President of The University of British Columbia* (1991), the court concluded (38) "that the instrument was not perfect, that it had flaws, and that the very limited number of samples (because of the very limited number of courses and students surveyed over the period) impaired its reliability. (p.30). (39) "However, we accept the evidence of Dr. [X] that the instrument has some value, directed toward the specified factors. The court noted that (28) "One problem with the questionnaire is that it solicits bad points as well as good points. Despite that caveat, we conclude that the inclusion of the qualitative comments was not a significant error" (p.32).
>
> In *Robert Kramer v. The President of the University of British Columbia* (1992), the Board said, (19) Given certain Departmental procedures, "there is a danger that some negative class commentary will dominate the discussion and will not be the 'independent' opinion of all of the students. (20) This is especially true in the context of the direction to assess "effectiveness" versus "popularity" (p.10). They further noted, (18) Given that "There was no peer review at all; no member of the Department audited any of Dr. Kramer's lectures. There was, therefore, nothing to guide the Department but the student comments," and "no way to test the accuracy or fairness of the undoubtedly disturbing comments in Asian Studies" (p.10).
>
> In *University of Regina Faculty Association v. University of Regina* (1993), The Board argued (6) that "the University was under an obligation to verify negative comments before acting on them" (p.4).
>
> In *Christopher Turner v. The President of the University of British Columbia* (1993), the Board said, (7) "while not ignoring some student unhappiness with Dr. Turner's teaching style, we think that the comments and emphasis on the size of Dr. Turner's classes as evidence of poor teaching are open to objection and constitute errors of procedure and/or evidence" (p.6).

Summary: With regard to establishing the general and statistical accuracy of SEF in the above cases, the range of opinion is from (14) statistical analyses may be a part of a plaintiff's effort to establish discriminatory treatment if it reached proportions comparable to those in cases establishing a prima facie racial discrimination, (15) cautioning that statistics are not irrefutable, with their usefulness depending the surrounding facts and circumstances of a case, (16) the Court maintaining that it need not consider and is under no obligation to establish the accuracy of administrative interpretations of SEF, (17) that tenure criteria are not drawn with "mathematical nicety," (18) administrator's failure to perform statistical comparisons is not arbitrary and is reasonable, (19) especially if such is not required by a Faculty Association Contract, (20) nearly any use made of SEF is acceptable if it followed the standard practice of the university, (21) that creativity, rapport with students and colleagues, teaching ability, and other qualities are intangibles which cannot be measured by objective standards.[18]

Numerical Ranking of Faculty

It seems to be common practice to rank and compare faculty to each other according to average SEF scores, considering decimal differences between faculty as significant.

> In *Dyson v. Lavery* (1976), a student evaluation ranked her 46th of 48 teachers.
>
> In *Lieberman v. Grant* (1979), the court noted (4) a compilation of student ratings showed that the cumulative ratings for members of the department ranged from a low of 4.09 to a high of 8.95. She had a cumulative rating of 7.06, which ranked her 12th out of the 15 junior faculty members. The 7.06 figure included the ratings from a previous semester in which the plaintiff received a rating of 8.18. Prior to this rating in the spring of 1972, the plaintiff's cumulative rating was 6.7.
>
> In *Carley v. Arizona Board of Regents* (1987), it was noted that (1) of the 13 faculty in his department of art, he was ranked fifth, (2) by his chairman he was ranked 7th, (3) student evaluations, however, ranked him last: 13th of 13 (p.1105).
>
> In *Robert Kramer v. The President of the University of British Columbia* (1992), the court noted (24) scores in the other two courses were higher---3.45 in one, 3.91 in another, against a "faculty average" of 4.22. The board further noted, "In the result, one got a 2.82 and one got a 3.07...the difference is statistically invalid in any event" (p.10).

Summary: Numerical scores often result in faculty (22) being compared relative to other faculty, (23) being ranked relative other faculty, (24) with distinctions often being made on the basis of tenths of a decimal, (25) with most courts accepting these fine decimal distinctions.

Rulings on The Qualitative and Subjective Assessment of SEF

Acceptance of Administrative Subjective Judgements of SEF Data

An issue directly related to both the reliance on and validity of SEF is views of the court regarding accepting or not accepting subjective administrative judgements of SEF.

> In *Dyson v. Lavery* (1976), the court found that despite questionable errors it concluded that administrative judgements were acceptable because, "they were sincere and grounded on some evidentiary basis" (p.111); and (5) "In the absence of a finding that same were sexually motivated, the administration's professional judgment must be respected" (p.111).
>
> In *William Sypher v. Vermont State Colleges Faculty Federation* (1982), (7) sufficient evidence exists from which the Dean and President could have reasonably concluded Sypher was not above average in his teaching effectiveness; (8) the Board went on to say that if they adopted the Colleges' view that Sypher was not reappointed because of his teaching effectiveness, no argument advanced by him defending his teaching was likely to persuade the President because his decision was made on the "vigor and variety of student criticisms" (p.135).
>
> In *Carley v. Arizona Board of Regents* (1987), The court ruled (18) the University president was free to consider factual findings made by minority members of the academic freedom and tenure committee and any other evidence which he found relevant in determining whether to deny renewal of teaching contract to non tenured instructor. The president was not bound by factual findings made by majority members of committee (p.1103).

> In *Yu Chuen Wei v. Vermont State Colleges Faculty Federation* (1995), it was noted that (28) The Dean and the President obviously had much experience in reviewing student evaluations, and could reasonably draw on that experience in each tenure review. (p.311); judgements "were not arbitrary or capricious and were exercised honestly upon due consideration,"....that Deans and Presidents have "much experience in reviewing student evaluations, and could reasonably draw on that experience" (p.311).
>
> In *Dr. Brian Maclean v. President of The University of British Columbia* (1991), the court said. (40) "The relevance and quality of the scores are "a matter of weight for the various decision-makers, and we assume that they were reasonably aware of the limitations of student evaluations and gave them the weight they deserve" (p.30).
>
> In *Robert Kramer v. The President of the University of British Columbia* (1992), the board concluded, "In the final analysis, we feel that this review of the Head's comments on teaching, which would be the sole evidence upon which the Dean and the President could rely, shows that it was incomplete and might have been misleading" (p.12-14).
>
> In *University of Regina Faculty Association v. University of Regina* (1993), he Board said teaching was wrongfully evaluated, but upheld denial of tenure on grounds of inadequate scholarship.
>
> In *Christopher Turner v. The President of the University of British Columbia* (1993), The board concluded that (11) "there were sufficient errors of procedure and/or evidence to return the case for reconsideration" (p.11).

Summary: Courts have tended to accept administrative subjective judgements if (26) they are deemed sincere (27) grounded on some evidentiary basis (28) if made on the "vigor and variety of student criticisms" (29) "not arbitrary or capricious and were exercised honestly upon due consideration," (30) based upon "much experience in reviewing student evaluations, (31) reasonably draw on that experience (32) and have ruled that Presidents are not bound by factual findings made by majority members of a faculty.

Use of Qualitative Written Student Comments

Over and above quantitative data, the use of written comments, often single instances, by students on their SEF seems wide spread by both educational administrators, faculty evaluation committees, and the courts.

> In *Dyson v. Lavery* (1976), the course said (1) "A number of students apparently had voiced displeasure over the quality of her class preparation and presentation" (p. 111 (3) "These *impressions*" said the court, "were largely confirmed after the initial decision to not rehire her had been made, by a student evaluation that ranked her 46th of 48 teachers in the Business Department" (p.111, italics added).
>
> In *Johnson v. University of Pittsburgh* (1977), the court said, (3) "we have the *instance* referred to in Finding 27 (p.1359, italics added).
>
> In *Lieberman v. Grant* (1979), the court noted (3) based on complaints received from "*several students*," to the effect that Lieberman's interest in feminism caused her to ignore other themes in literature (p.873, italics added).
>
> In *William Sypher v. Vermont State Colleges Faculty Federation* (1982), (1) some of the student comments noted that, "When students try to disagree he shoots you down and tries to degrade you in front of the class," (p.115), while others said, "encourages

student participation as much as possible... encourages student to express their ideas freely and not worrying how 'dumb' it may sound...always wants you point of view." (P.115) (2) With regard to the numerical ratings, the Board's opinion was that (3) "regardless of a strong majority of students' rating his teaching as above average, (4) the existence of a significant minority of students feeling degraded, humiliated, and embarrassed can reasonably lead an evaluator to question a teacher's effectiveness" (p.115).

In *Yu Chuen Wei v. Vermont State Colleges Faculty Federation* (1995), the Board said, (22) "the statistical comparison does not take account of the comments made by students on the evaluation forms. Grievant's student evaluations are striking in how often mention is made of Grievant's communication difficulties, particularly language difficulties (p.304-5). The board further noted with respect to comments that while some students had written that she was a "slant eyed bitch," and that she should "go back to China," (30) "We also are not persuaded that the racism evident in the student evaluations of Grievant made student evaluation results unreliable. The percentage of evaluations in which racism by students was evident was approximately one percent of the total evaluations" (p.306).

In *Robert Kramer v. The President of the University of British Columbia* (1992), (2). The department Head viewed Kramer's 1989-90 course evaluations "with some alarm"....(4) Even more disturbing to the department Head was that a considerable number of students in their written comments stated that Dr. Kramer was biased sarcastic, and hostile to the material and that a number of students had stated that Dr. Kramer's teaching would cause them to stay away from the Asian Studies department. (5) There were also some diametrically apposed positive comments" (p.10).

In *University of Regina Faculty Association v. University of Regina* (1993), The Board argued (6) that the University was under an obligation to verify negative comments before acting on them. Consequently, (7) the fact that Dr Jalan had received some negative evaluations from students could not be used to undermine the otherwise generally favorable comments he had received in his annual performance reviews" (p.4).

In *Dr. Brian Maclean v. President of The University of British Columbia* (1991), the court noted that (25) "With respect to the "qualitative" scores---i.e., the "comments," there was a clear error. The qualitative comments from a number of courses were read and commented on, and conclusions were drawn from them which went into the "file." Both Reviewing faculty read and commented on them, as did the Department Chair in her letter to the Dean. Yet the Dean had clearly stated in a departmental memo that the qualitative comments were not to be used for administrative or promotion purposes. (26) While in the abstract there is no reason why such comments would not be relevant, if the Department had a rule against their use, or in other words if they were "for the professor's eyes only," then it was a significant breach of Departmental rules to use them" (p.31). (27) In the opinion of the Board, so long as the comments were fairly presented, they offered the PAT [Promotion and Tenure Committee] and others a better balanced view of the teaching qualities and problems of Dr. MacLean than the quantitative statements alone" (p.31). (28) The court noted that "One problem with the questionnaire is that it solicits bad points as well as good points. Despite that caveat, we conclude that the inclusion of the qualitative comments was not a significant error" (p.32).

Summary: The use of student comments ranges from (33) placing importance on a single comment (34) to several comments as significant information, (35) maintaining that statistical analyses of SEF need to be bolstered by individual comments, (36) maintaining that while some very negative---e.g., racist, sexist---comments may be found, the court may find that they do not rendered the SEF unreliable, (18) that such instances or "impressions" may be validated after the fact, (37) negative comments often seem to outweigh positive ones, and (38) comments may often outweigh numerical data to the contrary, (39) negative comments need not be verified before acting on them, (40), that negative comments can not be used to undermine otherwise generally favorable comments received in an annual performance review.

Mixed Student Comments

Just as quantitative SEF data may be bimodal, so too written student comments may also be bimodal or mixed. How do courts view and pronounce on such data?

> In *Johnson v. University of Pittsburgh* (1977), the court noted (2) they "approached this question of teaching ability with considerable doubt, in view of the fact that in prior years there does not appear to have been any criticism of her teaching and also in view of the fact that...there was evidence that the department chairman, had informed her after one of her lectures in 1971 what a great lecture it had been"; On the other hand, the court said (3) "we have the *instance* referred to in Finding 27 (p.1359, italics added).
>
> In *Fields V. Clark University* (1987, it was observed (3) a few of which, from students in Fields' seminars, were "wildly enthusiastic" about her enthusiasm, commitment and presentations; (4) a few were ambivalent; (5) with a considerable number being extremely negative, particularly (6) with regard to her large lecture classes in basic courses in sociology.
>
> In *Yu Chuen Wei v. Vermont State Colleges Faculty Federation* (1995), moreover, they said, (19) "The statistical comparison demonstrates that Grievant was evaluated higher by students than [her male colleague] with respect to upper level classes, but that (20) [the male colleague] was evaluated higher than Grievant in lower level classes. Given (21) this "mixed" result, the statistical comparison of evaluations does not demonstrate by a preponderance of the evidence that Grievant's students rated her the same, or better, than [male colleague]" (p.305).
>
> In *Dr. Brian Maclean v. President of The University of British Columbia* (1991), it was noted that (20) In general, the in-class peer reports were mixed but favourable. The in-class discussions were more problematic. (p.30). (21) While the knowledge, interest and enthusiasm of Dr. MacLean were acknowledged, "the problem appeared to be one of style or personality." It was further noted that (29) "As against the low figures, they disclosed a number of good qualities in Dr. MacLean---enthusiasm for his subject, wide knowledge of the literature, much out of class assistance to students, and a commitment to seeking good work from students. (p.31). (30) The reviewing faculty report noted the comments about Dr. MacLean's "derogatory manner, biased opinion, unwillingness to listen," were matched by "clear, stimulating, very helpful after class." And, (31) "some students have told us that the comments made were not representative of the class as a whole and were unduly influenced by the process" (p.41). (32) "A number of students, both from earlier years and from his current classes, furnished letters of support, and in preparation for the appeal, some furnished affidavits with respect to particular matters such as the 'intimidation' discussion in Soc. 250 and events in Soc. 490 and 520 in the fall of 1989." (p.33)
>
> In *Robert Kramer v. The President of the University of British Columbia* (1992). (16) While a number of negative student comments were quoted in the department Head's letter, there were a number of very positive comments, and these were not mentioned at all. (25) "We have examined all of these written comments. There was a very wide range of comments. There were not 29 comments saying sarcastic and biased comments; but there were certainly 29 comments which included either cynical, sarcastic, biased, insulting, negative, condescending, belittling, opinionated, arrogant, nihilist, and destructive.... (29) However, it would only be fair to add that there were a number of comments in favour of Dr. Kramer, stating that the student "liked the

> course immensely," "now interested in Asian Studies"; "helps create a relaxed atmosphere," "really enjoyed him," "very approachable and knowledgeable," "very enthusiastic," "captivates audiences with his humour," "very effective" (p.12). (30) "In the other two courses, both small, both Japanese language, there were also some negative comments" (p.12).
>
> In *Christopher Turner v. The President of the University of British Columbia* (1993), the board noted that (6) "While there is no question of Dr. Turner's competence as a teacher at all levels, teaching evaluations for the last several years show that his effectiveness is marred by what students perceive as excessive formality, lack of enthusiasm and dullness....In a previous promotion attempt, his teaching was briefly described as "very competent" but student evaluations indicate further improvement to be "better than adequate" (p.2).

Summary: With regard to non numerically assessed written student comments, they are often qualitatively characterized as (41) a few were ambivalent, (42) a considerable number, (43) of mixed result, and selectively recognized: (44) it would only be fair to add that there were a number of comments in favor, (45) there were also some negative comments, (45) sometimes placing the greater weight on past evaluations of teaching over current comments, (47) sometimes placing greater weight on current comments over past positive evaluation of teaching.

Variables Not Taken Into Account When Assessing SEF

Student Biases in Affecting SEF

A significant issue is how courts view variables in assessing the reliability and validity of SEF is student bias affecting their evaluation of faculty.

> In *Johnson v. University of Pittsburgh* (1977), the court noted that (10) "It has also been pointed out that in some cases difficult courses have to be given to the students and the material is such that it is difficult for even the best teacher to get it across.
>
> In *Carley v. Arizona Board of Regents* (1987), he (7) characterized his professional style as being a "demanding teacher contrary to some student expectations," (8) Because of this, he maintained his popularity suffered and resulted in low student evaluations, (9) examination of his student comments indicated that Carley was correct in his assessment as 61% (49 out of 80) negative student comments focused on these values. The court ignored these findings.
>
> In *Dr. Brian Maclean v. President of The University of British Columbia* (1991), it was noted that (21) While the knowledge, interest and enthusiasm of Dr. MacLean were acknowledged, "the problem appeared to be one of style or personality."
>
> In *Robert Kramer v. The President of the University of British Columbia* (1992), the Board noted that (26) It was obvious that almost all of the classes were upset about an examination which was considered more geography than Asian Studies, and (27) they didn't like the marking. (28) They also felt the workload was far too heavy for an "introductory" course. The Board apparently only noted this variable.

Summary: Student bias variables include (48) being a demanding teacher, (49) thus thwarting student expectations, (50) difficult examinations (51) grading, (52) heavy workload in a course. (53) While most courts ignore these student biases in

SEF, (54) occasionally a court will recognize that difficult courses have to be given to the students and that such material is difficult for even the best teacher to get the material across.

Other Variables

Cases concerned with the validity of SEF note various factors that were not controlled in the evaluation of the faculty process. Courts sometimes weigh these factors heavily, sometimes they do not.

> In *Lieberman v. Grant* (1979), Lieberman attempted to introduce approximately ten personnel files concerning the tenure proceedings of other faculty in the English department for comparison. (6) Recognizing that such evidence would have had some minimal probative value, the Court, exercised its discretion under Fed. R.Ev. 403, and excluded it on the ground that "such probative value would be substantially outweighed by the delay and waste of time, which introduction of such evidence would have necessarily entailed....The plaintiffs case without such evidence seemed almost interminable, consuming 52 trial days over a two-year period. That is long enough" (p.873).
>
> In *Fields V. Clark University* (1987) the court notes but does not admonish the non separation of student remarks from small seminar courses and those from large lecture classes.
>
> In *Cynthia J. Fisher v. Vassar College* (1995), the district court found (2) that the biology department distorted Fisher's teaching recommendations by (3) "selectively exclud[ing] favorable ratings," by "focus[ing] on the two courses in which Dr. Fisher had difficulties" and (4) by "applying different standards to her than were applied to other tenure candidates" (p.1209).
>
> In *Yu Chuen Wei v. Vermont State Colleges Faculty Federation* (1995), it was noted that (19) "The statistical comparison demonstrates that Grievant was evaluated higher by students than her [male colleague] with respect to upper level classes, but that (20) [male colleague] was evaluated higher than Grievant in lower level classes." Given (21) this "mixed" result, the statistical comparison of evaluations does not demonstrate by a preponderance of the evidence that Grievant's students rated her the same, or better, than [male colleague]" (p.305). Wei maintained that (16) her students rated her the same or higher than the male colleague's students rated him. The Board disagreed, saying, (19) "We note that the comparison offered by Grievant is somewhat weak since [male colleague] was tenured in 1988, and those student evaluations of his which were compared with Grievant post-dated his tenure review by a number of years"...further saying, "we decline to hold such an involved comparison is necessary before a reasonable tenure determination can be made" (p.305).
>
> In *Dr. Brian Maclean v. President of The University of British Columbia* (1991), the Board noted that (19) the reviewing faculty held in-class discussions about his teaching.
>
> In *Robert Kramer v. The President of the University of British Columbia* (1992), Kramer argued that the most significant mistake was the failure to consider all aspects of his teaching. For example, only his teaching in 1989-90 was considered, whereas (9) he had taught a wide range of courses over the previous three years (10) had three new courses that year, (11) plus a graduate course. Moreover, (17) The department head indicated that his teaching was not up to the departmental "standard." The standard appeared to be the performance of the tenure-track faculty, though Kramer was one of the most junior faculty members (p.8). (15) Only one of the more than thirty numerically rated questions was used: "Rate instructor bad to good." (16) While a number of negative student comments were quoted in the department Head's letter,

> there were a number of very positive comments, and these were not mentioned at all.
>
> In *Christopher Turner v. The President of the University of British Columbia* (1993), the Dean said, "there were few students in undergraduate literature courses since 1986/7---(3,8, and 6 respectively," thus mistaking student 'response' figures for actual student enrolment. The Board concluded that (5) "This misunderstanding is in our opinion sufficient in itself for a reconsideration, since teaching was the focus..." (p.3), and (7) "we think that the comments and emphasis on the size of Dr. Turner's classes as evidence of poor teaching are open to objection and constitute errors of procedure and/or evidence" (p.6).

Summary: The variables noted in these cases include, (55) not controlling for class size, i.e., those obtained in small seminars from those obtained in large lecture classes, (56) those obtained from tenured faculty from those obtained from non tenured junior faculty, (57) not performing appropriate comparisons of SEF with other faculty, (58) noting SEF in all courses, not just to problem courses, (59) not mistaking student 'response' figures for actual student enrolment figures when using them to determine student attraction to a course, (60) using all courses taught, (61) taking into consideration the faculty teaching a wide range of courses, versus those with lighter teaching loads, (62) number of new courses taught in a year, (63) whether graduate courses were taught at the same time as teaching undergraduate courses, (64) selectively mentioning only negative student comments, or (65) overly weighting negative comments, and (66) different procedures for gathering student opinion.

Courts Impact On Academic Matters

Impact on Academic Standards

It has historically been the case that faculty as a collective body and the institution set the expected level of student performance.

> In *William Sypher v. Vermont State Colleges Faculty Federation* (1982), it was observed (9) With regard to the "political" aspect of the case, Sypher had written a letter in defense of his student rating level at the college which said, "it is certainly distressing when very good is not good enough, especially at a college with a modestly-talented student body that often discourages efforts at subtlety, wit and deeper penetration of subjects." (10) The Board responded to this letter saying, "other actions and statements by Grievant constituted legitimate reasons for not retaining him. In a May, 1980, letter to, Dean Beston, Grievant expressed his contempt for Castleton students" (p.135), (11) concluding, "Accordingly, we find credible the College's contention that Grievant was not reappointed because of his teaching effectiveness. [Italics added] (p.135).V.
>
> In *Lovelace v. Southeastern Massachusetts University* (1986), the court said, (4) A school sets itself up to attract and serve only the best and the brightest students or whether it instead gears its standard to a broader, more average population is a policy decision which, we think universities must be allowed to set....matters such as course content, homework load, and grading policy are core university concerns (p.424).[19]
>
> In *Robert Kramer v. The President of the University of British Columbia* (1992), the Board said (30) "One perceptive student noted that some of the unhappiness came from the fact that the levels of Japanese language ability were badly divided; some found it easy, others very hard" (p.12).

Summary: From the above limited cases, the courts have clearly said (67) universities must be allowed to set (68) standards, including (69) course content, (70) homework load, and (71) grading policy.

Methods of Instruction and Academic Freedom

It is generally accepted that academic freedom clearly pertains to speech in the classroom. In like manner, most faculty seem to believe that academic freedom also pertains also to how a course is taught. How do courts view teaching method in relation to SEF an academic freedom?

> In *Lovelace v. Southeastern Massachusetts University* (1986), the court's ruled, (1) "It is important to note what plaintiffs first amendment claim is and to separate speech from action. Plaintiff has not contended that he was retaliated against simply because he advocated that the university elevate its standards.... Plaintiffs complaint instead is that he was retaliated against when he refused to change his standards" (p.425), (2) citing other cases, the court rejected his contention that university teacher has first amendment right to disregard established curriculum content, that the first amendment does not prevent a university from terminating an untenured faculty whose pedagogical style and philosophy does not conform to those of the school's administration, (3) "is a policy decision which, we think, universities must be allowed to set" (p.426). Further, the court ruled that (4) "We will assume for purposes of this opinion that plaintiffs refusal to lower his standards was a substantial motivating factor, (see Mount Health Board of Education v. Doyle, 429 U.S. 274, 283-284, 97 S.Ct. 568, 574-575,50 LEd.2d 471 (1977), in the decision not to renew his contract.
>
> In *Carley v. Arizona Board of Regents* (1987), (4) he claimed as "protected speech" his teaching methods where his goal in his commercial art course was to promote a business atmosphere by requiring attendance, promptness, and self-reliance, and required them to meet deadlines. The court ruled, (19) his teaching style is not a form of speech protected under the First Amendment. (20) Decision not to retain a non tenured instructor, even if based, in part, upon student evaluations expressing disapproval of his teaching methods, did not violate instructor's First Amendment right to academic freedom; (21) Carley was not denied a contract because of expressing unpopular opinions or otherwise presenting controversial ideas to his students. (22) Thus, we conclude that the decision not to retain Carley, even if based, in part, upon student evaluations expressing disapproval of his teaching methods, did not violate his first amendment rights. (p.1103).

Summary: Teaching method (72) because it is an action, not speech, (73) is not a form of free speech, nor (74) covered under academic freedom, (75) except if noted in specific contractual faculty agreements.

Release of SEF To Students and To the Public

In addition to the administrative use of SEF, in recent years, other uses of SEF have become controversial, including releasing SEF to students and to the public.

> *University of Idaho*: A legal ruling, cited on the World Wide Web site of the *Topical Interest Group: Assessment in Higher Education* (Evaluating teacher evaluations, 1996), [20] notes that the University of Idaho also recently went to court over the issue of whether SEF can be published. The student newspaper initiated a lawsuit when it was refused access to SEF for publication. The legal question was whether SEF is

> protected under privacy rights by the Idaho Code. In a ruling that seems to strain logical credulity the court ruled that since the University did not consider students as the general public, the University was not breaking the law by allowing students access to the evaluations. Further, the opinion of the court was that according to State law, teacher evaluations are not protected as part of personnel records.
>
> *University of Wisconsin*: In another ruling, the Chancellor at the University of Wisconsin refused to release SEF, citing a statute that disallows personnel evaluations from being released to public view. Students took the chancellor to court. However, after being advised by the state's Attorney General, citing Wisconsin's open-records law, the University of Wisconsin's campus released SEF to the public. Both the student and faculty senates, passed resolutions in support of the Chancellor's refusal, and the university's lawyer concurred. Despite these resolutions, the Attorney General disagreed, writing that "the requested records are public records and the university's stated reasons for withholding access do not outweigh the public interest in the records" (Chronicle of Higher Education, 1994a, 1994b).

Summary: (76) Unlike most personnel records, SEF can be released to students and the public, on the grounds, that (77) students are not considered the general public, and (78) that SEF are public records and withholding them from public access does not outweigh the public interest in them.

Conclusion

From the above cases, it is clear that "a view from the courts" on SEF does not mean a coherent set of rulings. Given the lack of legal attention to SEF, this is not surprising. A collectively incoherent set of rulings, however, is not restricted to the concept of SEF but for the time-honored concept of academic freedom as well. As Kaplin and Lee (1995) conclude,

> The foregoing discussions have made clear that academic freedom is an area in which the law provides no firm guidelines for administrators. This is particularly true for private institutions, since the decided cases are almost all constitutional decisions applicable only to public schools. Even the constitutional cases are sometimes incompletely reasoned or difficult to reconcile with one another. The fact that decisions often depend heavily on a vague balancing of faculty and institutional interests in light of the peculiar facts of the case makes it difficult to generalize from one case to another (p.192).

When we speak of "the courts," then, we are speaking of a hydra with some common but many disparate voices.

Given the above rulings and the courts propensity to accept faculty/institutional agreements, it would seem clear as Kaplin and Lee advise, that, "it is especially crucial for institutions to develop their own guidelines on academic freedom and to have internal systems for protecting academic freedom in accordance with institutional policy" (p. 192). This would appear to be especially true for SEF policy.

From most of the above cases---even given that, as challengers, the burden of proof has been on faculty --- it seems clear that the courts have not been kind to faculty with regard to student evaluations.[21] It is the burden of the two following papers to explore, analyze, and comment on the implications of the views from the court presented here.

Notes

1. Address correspondence to: Robert E. Haskell, Ph.D. Professor of Psychology, Department of Social and Behavioral Sciences, University of New England, Biddeford, ME 04005. Email: haskellr@gwi.net. I would like to thank Mr. Nicholas DiGiovanni, Jr., attorney at law, of Morgan, Brown & Joy, Boston, MA, Mr. Gary Founds of College Legal Information, Inc., especially Professors William A. Kaplin, School of Law, Catholic University of America, and John Damron, of Douglas College for continually providing me with sources and support.
[BACK to document]

2. The issue of needing some form of faculty performance evaluation is an immediate and pragmatic one that can not be dealt with here. If used correctly (see Copeland and Murry, 1996; Kemp and Kuman, 1990; Scriven, 1995, 1993, 1991,1988; Seldin, 1984), SEF can be very useful instructionally, and when used in conjunction with other methodologically sound evaluation procedures and criteria, it can assist in informing an institution when a faculty may be having problems with his or her teaching and/or relations with students.
[BACK to document]

3. In addition to there having been no categorical mention on the professional level of SEF as an infringement on academic freedom, the reasons for SEF not being examined as an infringement on academic freedom and tenure rights are that many faculty (1) are often embarrassed to admit that student evaluations may influence their professional behavior in the classroom, (2), seem to accept that to question the right of students to evaluate faculty is considered unprofessional if not undemocratic, (3) seem also to accept that to question the right of students to evaluate faculty is seen as self serving, if not downright mean spirited, and (4) see research on such SEF as not high status research.
Based on a review of statements in the literature, in a previous paper I have, outlined faculty attitudes on SEF impinging on tenure and academic freedom (Haskell, 1997). Statements by faculty contend that SEF (1) is prime facie evidence of administrative intrusion into the classroom, (2) are often used as an instrument of intimidation forcing conformity to selected standards (Young, 1993), (3) create pressure for lowered teaching standard (Bonetti, 1994), (4) are responsible for a considerable amount of grade inflation (Greenwald, 1997), (5) function as prescriptions for classroom demeanor (Damron, 1996), (6) when used for promotions, salary raises or continued employment, SEF becomes a potent means of manipulating the behavior of faculty (Stone, 1995), (7) when salary and promotion are possible consequences of SEF there is pressure for faculty to teach in a manner that results in higher student evaluations (Damron, 1996), (8) contrary to their original intent of improving instruction, do not eliminate poor or below-average instructors but instead increases poor teaching practices (Carey, 1993; Barnett,1996), (9) illustrate a mercantile philosophy of "consumerism" (Benson, and Lewis, 1994), which erodes academic standards (Goldman, 1993; Renner, 1981), (10) have thus lowered the quality of U.S. education (Carey, 1993; Crumbley, and Fliedner, 1995; Young, 1993), (11) lead to the inappropriate dismissal of faculty (Parini, 1995), and (12) constitute a threat to academic

freedom (Dershowitz, 1994). Finally, SEF creates an educational conflict of interest between faculty and student that negatively impacts on the quality of instruction. [See also Furedy, 1995.] [BACK to document]

4. Based on a review of statements in the literature, in a previous paper I have, outlined faculty attitudes on SEF impinging on tenure and academic freedom (Haskell, 1997). Statements by faculty contend that SEF (1) is *prime facie* evidence of administrative intrusion into the classroom, (2) are often used as an instrument of intimidation forcing conformity to selected standards (Young, 1993), (3) create pressure for lowered teaching standard (Bonetti, 1994), (4) are responsible for a considerable amount of grade inflation (Greenwald, 1997), (5) function as prescriptions for classroom demeanor (Damron, 1996), (6) when used for promotions, salary raises or continued employment, SEF becomes a potent means of manipulating the behavior of faculty (Stone, 1995), (7) when salary and promotion are possible consequences of SEF there is pressure for faculty to teach in a manner that results in higher student evaluations (Damron, 1996), (8) contrary to their original intent of improving instruction, do not eliminate poor or below-average instructors but instead increases poor teaching practices (Carey, 1993; Barnett,1996), (9) illustrate a mercantile philosophy of "consumerism" (Benson, and Lewis, 1994), which erodes academic standards (Goldman, 1993; Renner, 1981), (10) have thus lowered the quality of U.S. education (Carey, 1993; Crumbley, and Fliedner, 1995; Young, 1993), (11) lead to the inappropriate dismissal of faculty (Parini, 1995), and (12) constitute a threat to academic freedom (Dershowitz, 1994). Finally, SEF creates an educational conflict of interest between faculty and student that negatively impacts on the quality of instruction.
[BACK to document]

5. Increasingly, SEF are becoming a primary and single factor in denial cases.
[BACK to document]

6. I had a Lexis Nexus search conducted but with only a couple of usable cases found. I have been in contact with a number of legal scholars who have written compendia of legal cases regarding the denial of tenure and promotion. Few were able to refer me to cases involving SEF.
[BACK to document]

7. The term "court" as used here will also include rulings by state Arbitration Boards. Since this paper is only concerned with how ruling bodies view SEF data, this paper will not distinguish between state courts, federal courts or arbitration boards.
[BACK to document]

8. Because the concept of context in itself is complex, it would take this paper too far afield to attempt to explicate why certain case data were included while other omitted.
[BACK to document]

9. I will explore the consequences of this issue for SEF my third paper.
[BACK to document]

10. The weight given to SEF of faculty vary according to the institution, with research oriented institutions not according them the weight that teaching oriented institutions do. The latter, of course, constitute the vast majority of campuses.
[BACK to document]

11. The question, of course is: does student feedback to faculty result in improved teaching and student learning. In a review of studies, Marsh (1984) suggests that there is a small positive correlation for improved student learning if SEF feedback to faculty if used in a carefully constructed collegial consultation process. A condition which seldom occurs.
[BACK to document]

12. In the appendix, I have provided an abstracting of each case relative to its SEF content. What is needed are legal scholars to locate and review a larger data set and to perform a statistical content and comparative analysis of court rulings on SEF relative to different contextual variables.
[BACK to document]

13. While there are certainly insufficient cases cited here to draw any firm conclusions, it does appear that at least in relations to SEF that Canadian Boards seem to take a more critical look at their use than do courts in the U.S. As I understand it, arbitrations are adjudicated by a panel made up of a arbitrator who is a lawyer and is officially recognized by the government; typically, two other lawyers are selected by the university the Faculty Association. This procedure is circumscribed in provincial law and decisions rendered by the panel have legal standing. I would like to specifically thank Dr. John Damron of Douglas College for providing me with all of the Canadian cases with the assistance of the Canadian Association of University Teachers (CAUT).
[BACK to document]

14. While many educational issues in Canada and the U.S., are quite similar (see Emberley, P. C. & Newell, W.R. 1994; Emberley, P. C., 1995), it is well known that Canadian jurisprudence does not define free speech and academic freedom as broadly as does the U.S., system. While there are certainly insufficient cases cited here to draw any conclusions, it does appear that at least in relations to SEF that Canadian Boards seem to take a more critical look at their use than do courts in the U.S.
[BACK to document]

15. For reference convenience, the numbers in parentheses have been retained and correspond to the numbers in the abstracted appendices relative to the specific case.
[BACK to document]

16. The numbers in these brief summaries will be cumulative throughout the following summaries and a cumulative summary is presented in the appendix.
[BACK to document]

17. The problem, however, is assessing the validity of SEF data. At this time, state-of-the-art analyis of SEF Barnett, 1996; Greenwald, 1997; Greenwald & Gillmore (1996) for state of the art statistical analyses of previous studies suggesting their validity.
[BACK to document]

18. In general, courts have tended to only require precise accuracy in cases where EEOC issues are involved. I will address this issue in more depth in a following paper.
[BACK to document]

19. An important issue here is: who is the university? Faculty or administration. According to Kaplin (personal communication, April 1, 1997) "When courts seem

to equate the university administration or a single administrator with the "university," it is generally because the university's governing body has delegated the pertinent authority to the administration or administrator." I will be exploring the implications of this issue in more detail in Part IV.
[BACK to document]

20. See http://marsquadra.tamu.edu/TIG/GeneralArticles/evaluatingteacherevaluations.ht.
[BACK to document]

21. As one scholar (Damron, personal communication, April, 1997) who read a draft of this paper observed: from

> "the legal decisions you review in your paper it is clear that untenured and/or politically incorrect faculty are often considered to be "fair game" by administrators, with literally _any_ superficially plausibly excuse serving as a rationale for dismissal. Use of such strategies reveal that faculty are often regarded as little more than term employees who are as disposable (and replaceable) as tissues. Clearly, there is a very serious ethical issue here, and a hugely hostile attitude toward academic freedom and faculty in general....the great variety of decisions you've reviewed and their assorted implications for the coherence and ethics of the legal processes that gave rise to them...it seems to me that many judges and arbitration panelists have little sense of how to proceed in hearing[s] involving academics."

[BACK to document]

22. In general, courts have tended to only require precise accuracy in cases where EEOC issues are involved. I will address this issue in more depth in the fourth and final paper of this series.
[BACK to document]

References

Allegre, M., and Guista, M. (1993). What every English teacher ought to know about using student evaluations. *Teaching English in the Two-year College, 20*(1), 57-63.

Baez, B, and Centra, J. (1995). Tenure, promotion, and reappointment: Legal and administrative implications. *Ashe-eric Higher Education Report*, No. 1.

Barnett, L.D. (1996). Are teaching evaluation questionnaires valid? Assessing the evidence. *Journal of Collective Negotiations in the Public Sector, 25*(4), 335-349.

Benson, D.E., and Lewis, J.M. (1994). Students' evaluation of teaching and accountability: Implications from the Boyer and the ASA reports. *Teaching Sociology, 22*, 195-99.

Blum, D.E. (1990, October 3). U. of Guam removes evaluations from files. *The Chronicle of Higher Education*. Section: Personal & Professional, A21.

Bonetti, S. (1994). On the use of student questionnaires. *Higher Education Review, 26*, 57-64.

Brian Maclean v. President of The University of British Columbia (1991). Held at Vancouver, B.C. January 28--June 20.

Carey, G. W. (1993). Thoughts on the lesser evil: Student evaluations. *Perspectives on Political Science, 22*, 17-20.

Carley V. Arizona Board of Regents, 737 P.2d 1099 (Ariz. App. 1987).

Cashin, W. E. (1995). Student ratings of teaching: The research revisited. *IDEA Paper No. 32.* Manhattan, KS: Kansas State University, Center for Faculty Evaluation and Development.

Christopher Turner v. The President of the University of British Columbia (1993). Arbitration Appeals Board Case No.: AI-256 Vancouver, B.C. July 19,20 and 21.

The Chronicle of Higher Education. (1994a, April 27). State calls for release of faculty evaluations, p. A4

The Chronicle of Higher Education. (1994b, December 14). Faculty evaluations to go on line at u. of Texas at Austin, p. A21

Chuen Wei and the Vermont State Colleges Faculty Federation, 18 VLRB 261 (1995)

Copeland, J.D. and Murry, J. W., Jr. (1996). Getting tossed from the ivory tower. *Missouri Law Review, 61*, 233-327.

Crumbley, L.D., and Fliedner, E. (1995). *Accounting administrators' perceptions of student evaluation of teaching (SET) information*. Manuscript, Department of Accounting, Texas A&M University.

Cynthia J. Fisher v. Vassar College (1995). United States Court of Appeals for the Second Circuit Nos. 1179, 1303, 2275 Docket Nos. 94-7737, 94-7785, 94-9125

Damron, J.C. (1996). *Instructor personality and the politics of the classroom.* Manuscript, Douglas College, New Westminster, British Columbia, Canada V3L 5B2. (An Earlier Version in the June 1994 Issue of Faculty Matters (No. 5, Pages 9-12) and the September, 1994 Issue of Update (The Newsletter of the Okanagan University College Faculty Association). Available online at http://vax1.mankato.msus.edu/~pkbrandon/Damron_politics.html

Damron, (1997). Personal communication.

Dershowitz, A. (1994). *Contrary to popular opinion*. New York: Berkeley Books.

Dyson V. Lavery, 417 F.supp. 103 (E.d. Va. 1976).

Evaluating teacher evaluations. (1996). This article submitted by glen9579@novell@uidaho.edu on 11/12/96. http://marsquadra.tamu.edu/TIG/Forum.html. (See, Topical Interest

Group:Assessment in Higher Education at http://marsquadra.tamu.edu/TIG/GeneralArticles/evaluatingteacherevaluations.ht)

Fields V. Clark University, Civil Action No. 80-1011-s, United States District Court for the District of Massachusetts, 40 Fair Empl. Prac. Cas. (Bna) 670, March 14, 1986; Vacated and Remanded May 8, 1987.

Furedy, John J. (1995). Academic freedom, opinions and acts: The voltaire-mill perspective applied to current canadian cases University of New Brunswick Law Journal, 44, 131-4.

Goldman, L. (1993). On the erosion of education and the eroding foundations of teacher education. *Teacher Education Quarterly, 20* (Spring), 57-64.

Greenwald, A.G. (1996). *Applying social psychology to reveal a major (but correctable) flaw in student evaluations of teaching*. University of Washington, Draft Manuscript, March 1, submitted for publication.

Greenwald, A., and Gillmore, G. (1997). No pain, no gain? The importance of measuring course workload in student ratings of instructions. *Journal of Educational Psychology* (forthcoming).

Haskell, R.E. (1997). Academic freedom, tenure, and student evaluation of faculty: Galloping polls in the 21st century. *Education Policy Analysis Archives, 5*(6), [Refereed Online Journal] Available: http://olam.ed.asu.edu/epaa/v5n6.html

Husbands, C., and Fosh, P. (1993). Student's evaluation of teaching in higher education: Experiences from four European countries and some implication of the practice. *Assessment & Evaluation on Higher Education, 18*, 95-114

Johnson V. University of Pittsburgh, 435 F.supp. 1328 (W.d. Pa. 1977).

Kaplin, W.A. (1997). Personal communication. January 20th, School of Law, Catholic University of America.

Kaplin, W.A., and Lee, B. (1995). *The law of higher education: A comprehensive guide to legal implications of administrative decision making (3rd Ed)*. San Francisco: Jossey Bass.

Kemp, B.W., and Kuman, G.S. (1990). Student evaluations: Are we using them correctly? *Journal of Education for Business, 65*, (November/December). 106-11.

Lieberman V. Grant, 474 F.supp. 848 (D.conn. 1979).

Lovelace v. Southeastern Massachusetts University, 793 F.2d 419 (1st Cir.1986).

Marsh, H.W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology, 76*, 707-754.

Parini, J. (1995, July 14). Tenure and the loss of faculty talent. *The Chronicle of*

Higher Education, Section: Opinion, POV.

Peters V Middlebury College, 409 F.supp. 857 (D. Vt. 1977).

Poch, R.K. (1993). Academic freedom in American higher education: Rights, responsibilities, and limitations. *Ashe-eric Higher Education Report No. 4.*

Renner, R.R. (1981). Comparing professors: How student ratings contribute to the decline in quality of higher education. *Phi Delta Kappan, 63*, (2 Oct.), 128-30.

Robert Kramer v. The President of the University of British Columbia (1992). Arbitration Appeals Board Case No. AI-245. Held at the University April 22 and 23.

Scriven, Michael. (1988). The new crisis in teacher evaluation: The improper use of 'research-based' indicators. *Professional Personnel Evaluation News*. (January) p.4.

Scriven, M. (1991). *Evaluation Thesaurus*. (4th Ed.) Newbury Park, CA : Sage Publications.

Scriven, M. (1993). The validity of student ratings: In teacher evaluation. *Evaluation & Development Group*.

Scriven, M. (1995). A unified theory approach to teacher evaluation. *Studies in Educational Evaluation, 21*, 111-129

Seldin, P. (1993, July 21). When students rate professors. *The Chronicle of Higher Education*. Opinion p. POV.

Stone, J.E. (1995). Inflated grades, inflated enrollment, and inflated budgets: An analysis and call for review at the state level. *Education Policy Analysis Archives, 3*, Number 11 (2075) [Online] Available: http://olam.ed.asu.edu/epaa/v3n11.l:tml

University of Regina V. University of Regina Faculty Association and Dr. Pradeep Jalan (1993). Case No.: AI-298 Regina, Saskatchewan, July 9.

Weeks, K. M. (1996). *Managing departments: Chairpersons and the law.* Nashville, Tennessee: College Legal Information, Inc.

William Sypher v. Vermont State Colleges Faculty Federation, 5 VLRB 102 (1982).

Young, R. D. (1993). Student evaluation of faculty: A faculty perspective. *Perspectives on Political Science, 22*, 12-16.

Yunker, B.J., and Sterner, J (1988). A survey of faculty performance evaluation in accounting. *Accounting Educators' Journal*, (Fall). 63-74.

Zirkel, P.A. (1996). *The law of teacher evaluation*. Bloomington, In: Phi Delta Kappa Educational Foundation.

Appendix A: Summary of 78 Findings

From the cases reviewed, it can be seen that court rulings range from saying that (1) relying primarily or solely on student evaluations is acceptable, to (2) placing little exclusive reliance on SEF, (3) in rare cases SEF can not be permitted to stand in the way of promoting or retaining professors who are excellent in non teaching areas, (4) tenure decisions can not be based solely on SEF by students who have not been made aware of the ramifications of their evaluations, (5) anonymous documents or those "based on hearsay" should not be included in a faculty member's file, (6) students should be made aware of the purpose and ramifications of their evaluations of faculty, (7) anonymous student evaluations should not be used, (8) peer evaluations must also be a part of evaluating teaching, that (9) in cases of exceptional research faculty that popularity should not play a role in termination due to teaching, to (10) in normal cases that a measure of popularity is related to teaching effectiveness.

If (11) the courts and educational administrations can not allow low SEF to stand in the way of promoting or retaining professors who may be world renowned scientists, (12) deemed nationally or internationally exceptional as a researcher courts may disregard SEF, (13) but at least in these two cases they did not find the faculty exceptional. With regard to establishing the general and statistical accuracy of SEF in the above cases, the range of opinion is from (14) statistical analyses may be a part of a plaintiff's effort to establish discriminatory treatment if it reached proportions comparable to those in cases establishing a prima facie racial discrimination, (15) cautioning that statistics are not irrefutable, with their usefulness depending the surrounding facts and circumstances of a case, (16) the Court maintaining that it need not consider and is under no obligation to establish the accuracy of administrative interpretations of SEF, (17) that tenure criteria are not drawn with "mathematical nicety," (18) administrators failure to perform statistical comparisons is not arbitrary and is reasonable, (19) especially if such is not required by a Faculty Association Contract, (20) nearly any use made of SEF is acceptable if it followed the standard practice of the university, (21) that creativity, rapport with students and colleagues, teaching ability, and other qualities are intangibles which cannot be measured by objective standards.[22]

Numerical scores often result in faculty (22) being compared relative to other faculty, (23) being ranked relative other faculty, (24) with distinctions often being made on the basis of tenths of a decimal, (25) with most courts accepting these fine decimal distinctions.Courts have tended to accept administrative subjective judgements if (26) they are deemed sincere (27) grounded on some evidentiary basis (28) if made on the "vigor and variety of student criticisms" (29) "not arbitrary or capricious and were exercised honestly upon due consideration," (30) based upon "much experience in reviewing student evaluations, (31) reasonably draw on that experience (32) and have ruled that Presidents are not bound by factual findings made by majority members of a faculty. The use of student comments ranges from (33) placing importance on a single comment (34) to several comments as significant information, (35) maintaining that statistical analyses of SEF need to be bolstered by individual comments, (36) maintaining that while some very negative---e.g., racist, sexist---comments may be found, the court may find that they do not rendered the SEF unreliable, (18) that such instances or "impressions" may be validated after the fact, (37) negative comments often seem to outweigh positive one, and (38) may often outweigh numerical data to he contrary, (39) negative comments need not be verified before acting on them, (40), that negative comments can not be used to undermine otherwise generally favorable comments received in an annual performance review.

With regard to non numerically assessed written student comments, they are often qualitatively characterized as (41) a few were ambivalent, (42) a considerable number, (43) of mixed result, and selectively recognized: (44) it would only be fair to add that there were a number of comments in favor, (45) there were also some negative comments, (45) sometimes placing the greater weight on past evaluations of teaching over current comments, (47) sometimes placing greater weight on current comments over past positive evaluation of teaching. Student bias variables include (48) being a demanding teacher, (49) thus thwarting student expectations, (50) difficulty examinations (51) grading, (52) heavy workload in a course. (53) While most courts ignore these student biases in SEF, (54) occasionally a court will recognize that difficult courses have to be given to the students and that such material is difficult for even the best teacher to get the material across.

Some of the variables noted in cases include, (55) not controlling for class size, i.e., those obtained in small seminars from those obtained in large lecture classes, (56) those obtained from tenured faculty from those obtained from non tenured junior faculty, (57) not performing appropriate comparisons of SEF with other faculty, (58) noting SEF in all courses, not just to problem courses, (59) not mistaking student 'response' figures for actual student enrolment figures when using them to determine student attraction to a course, (60) using all courses taught, (61) taking into consideration the faculty teaching a wide range of courses, versus those with lighter teaching loads, (62) number of new courses taught in a year, (63) whether graduate courses were taught at the same time as teaching undergraduate courses, (64) selectively mentioning only negative student comments, or (65) overly weighting negative comments, and (66) different procedures for gathering student opinion. Courts sometimes weigh these factors heavily, sometimes they do not (67) universities must be allowed to set (68) standards, including (69) course content, (70) homework load, and (71) grading policy Teaching method (72) because it is an action, not speech, (73) is not a form of free speech, nor (74) covered under academic freedom, (75) except if noted in specific contractual faculty agreements, (76) since students may not be considered the general public students may have access SEF, (77) in some State law, teacher evaluations are not protected as part of personnel records. (78) SEF are public records because withholding access does not outweigh the public interest.

Appendix B: Summary of Case Outcomes

I. Dyson v. Lavery, (1976).

Patricia Dyson, a lawyer, in not renewing her contract, filed suit alleging sex and salary discrimination in both her initial employment and in the decision not to renew her contract. The Court concluded that Dyson's termination was not the product of sex discrimination.

II. Johnson v. University of Pittsburgh, (1977).
Sharon Johnson, at the end of her second three-year contract as an assistant professor was denied promote to the rank of associate professor and wan not granted tenure. She contended that her termination was based on sex discrimination. The University of Pittsburgh, claimed that the denial was a result of Johnson's inadequacy as a teacher and as a researcher and writer.

The court found that plaintiff's teaching was not adequate, upholding the university decision.

III. Peters v. Middlebury College, (1977).

Joan Peters file suit for sexual discrimination after being denied reappointment on the basis of her scholarship and teaching. Court found that she did not merit tenure, but was discriminated against.

IV. Lieberman v. Grant, 474 F.Supp. 848 (D.Conn. 1979).

Marcia R. Lieberman, denied tenure, claimed sex discrimination. The court concluded that sexual prejudice did not play any significant role in the tenure proceedings.

V. William Sypher v. Vermont State Colleges Faculty Federation (1982).

Sypher contented his denial of reappointment after his third year violated his academic freedom rights as stated in the faculty collective bargaining agreement. The faculty contended he was not reappointed because of his "political" activity on campus. The college contented he was not reappointed because of his inadequate teaching evaluations. The court found credible the College's contention that Grievant was not reappointed because of his teaching effectiveness.

VI. Lovelace v. Southeastern Massachusetts University, 793 F.2d 419 (1st Cir.1986).

Matthew Lovelace, a non tenured faculty challenged the decision not to renew his contract based on various constitutional claims, including a claim that the use of student ratings violated his academic freedom. He contended that his contract had not been renewed because he was a demanding teaching. When he refused to change his standards, his contract was not renewed. The court upheld the university.

VII. Fields V. Clark University (1987).

Fields contended she was being discriminated against on the basis of her sex. The university claimed she was denied reappointment because of inadequate teaching. The court ruled in Fields' favor.

VIII. Carley v. Arizona Board of Regents, (1987).

Denny Carley was denied reappointment largely on the basis of student evaluations of his teaching. This appeal is from a superior court judgment affirming a decision of Northern Arizona University President to deny the renewal of a teaching contract to Carley. The issues on appeal involved: (1) whether his right to academic freedom was violated because student evaluations were utilized as the primary tool to determine his teaching effectiveness. The court concluded that the decision not to retain Carley, even if based, in part, upon student evaluations expressing disapproval of his teaching methods, did not violate his first amendment rights.

IX. Robert Kramer v. The President of the University of British Columbia (1992).

Kramer, an Assistant Professor, was denied his first two-year reappointment. He maintained that there were errors of procedure, failures to obtain relevant evidence, and that the denial was based on a defective dossier. The Board upheld Kramer.

X. Christopher Turner v. The President of the University of British Columbia (1993).

Christopher Turnerwas denied promotion to full professor. The board concluded that (11) there were sufficient errors of procedure and/or evidence to return the case for reconsideration.

XI. University of Regina Faculty Association v. University of Regina (1993).

Pradeep Jalan, an Associate professor in his fifth year at the University. Denial of tenure was on the basis that he had not met the standard for an Associate Professor in the areas of scholarly performance and teaching. The Board ruled that the evidence relied upon by the employer to suggest that the standard was not met, in our view, was not sufficient.

XII. Cynthia J. Fisher v. Vassar College, (1995).

Cynthia Fisher was denied tenure. The college maintained the denial was because of her performance. Following a bench trial, the district court found that the college had discriminated against her by reason of her sex and age.

XIII. Yu Chuen Wei v. Vermont State Colleges Faculty Federation (1995).

Yu Chuen Wei was denied tenure. The case involved SEF as a part of a wider issue of racial discrimination. The Board concluded that Wei did not established discrimination. The Board concluded that Wei had not established that the tenure decision in her case was unreasonable, arbitrary or based on erroneous reasons.

XIV. Guam Federation of Teachers v. The University of Guam (1990).

XV. Dr. Brian Maclean v. President of The University of British Columbia (1991).

Denied reappointment after his first two-year term. The court upheld the university decision.

Appendix C: Abstracted Case Material

From each case, I abstracted out the material that was relevant to teaching effectiveness and SEF.

I. Dyson v. Lavery, (1976).

Patricia Dyson, a lawyer, hired to teach in the Business Department at Virginia Polytechnic Institute (VPI) in February, 1970, When VPI failed to renew her contract, she filed suit against the president of the university alleging sex and salary discrimination in both her initial employment and in the decision not to renew her contract. The court carried out a compete statistical analysis of salary and promotion by gender.

In terms of her teaching, the court said, "the Business Department Administrators testified that (1) the input received through normal channels for Ms. Dyson's teaching performance had been generally negative. A number of students apparently had voiced displeasure over the quality of her class preparation and presentation" (p. 111) and (2) she refused to follow normal procedures not posting student grades; (3) "These impressions" said the court, "were largely confirmed after the initial decision to not rehire her had been made, by a student evaluation that ranked her 46th of 48 teachers in the Business Department." Further, it was ruled that, (4) "The Court need not consider the accuracy of these administrative determinations, for it concludes that they were sincere and grounded on some evidentiary basis" (p.111); and (5) "In the absence of a finding that same were sexually motivated, the administration's professional judgment must be respected" (p.111 all italics added). The Court concluded that Dyson's termination was not the product of sex discrimination.

II. Johnson v. University of Pittsburgh (1977).

Sharon Johnson was an assistant professor at the University of Pittsburgh in the Biochemistry Department of the School of Medicine. At the end of her second three-year contract as an assistant professor, the university decided not to promote her to the rank of associate professor and not to grant her tenure. She contended that her termination was based on sex discrimination. The University of Pittsburgh, on the other hand, claimed that the denial of promotion and tenure was a result of Johnson's inadequacy as a teacher and as a researcher and writer.

The tenured faculty (1) described her teaching as "disastrous" and stated she demonstrated "a virtually complete inability to organize and convey a substantive body of information to students"; (2) the court noted they "approached this question of teaching ability with considerable doubt, in view of the fact that in prior years there does not appear to have been any criticism of her teaching and also in view of the fact that at...there was evidence that the department chairman, had informed her after one of her lectures in 1971 what a great lecture it had been"; On the other hand, the court said, (3) "we have the instance referred to in Finding 27 and the fact that in recent years increasing weight has been placed upon teaching ability"; (4) "We further have an evaluation by another faculty who impressed the court as completely impartial and credible who stated...that plaintiff's teaching performance was inadequate"; in addition, (5) "The court has placed little reliance on students' surveys. The value of these surveys was deprecated by the Dean and it was further testified that student surveys always result in the students in a given course rating a teacher, or professor, some of them as excellent, others as terrible and in between many who say passable, mediocre etc"; (6) "While the court is reluctant to rely upon student's surveys as demonstrating the fitness or unfitness of a faculty member nevertheless we cannot say it was unreasonable for the tenured faculty to consider this along with other matters in the meeting" (p.1359). (7) Citing a previous case (International

Brotherhood of Teamsters v. U. S, -- U.S. --, 97 S.Ct. 1843, 52 L.Ed.2d 396 (1977), where in a class action case the court cited McDonnell Douglas for the use of statistics) went on to say: "We have repeatedly approved the use of statistical proof where it reached proportions comparable to those in this case to establish a prima facie case of racial discrimination in jury selection cases . . Statistics are equally competent in proving employment discrimination. We caution only that statistics are not irrefutable. They come in an infinite variety and, like any other kind of evidence they may be rebutted. In short, their usefulness depends on all of the surrounding facts and circumstances" (8) The court further in Footnote # 20 said: "Considerations such as small sample size may of course detract from the value of such evidence" (p.1361); the court also noted, "It is obvious that a professor may be possessed of excellent qualifications as a research scientist and not necessarily be able to prove his or her worth as a teacher. It is also obvious that the court and the administration of universities cannot permit students to exercise a veto over professors who may be world renowned scientists and yet if the students rate them unfavorably can be terminated at any time because of unpopularity"; concluding that, (9) "in cases where one has an outstanding scientist of national or international reputation, one may decide to promote and give tenure notwithstanding inability to come across as a teacher, this however is not one of those cases. Finally, it said, (10) "It has also been pointed out that in some cases difficult courses have to be given to the students and the material is such that it is difficult for even the best teacher to get it across. (11) The court finds the weight of the evidence is that plaintiff's teaching was not adequate and while there are some allegations to the contrary and there had previously been no criticism of her performance, nevertheless the court cannot say that the tenured faculty was unreasonable in considering this factor and in arriving at the conclusion they did" (p.1366-7).

III. Peters v. Middlebury College, (1977).

Joan Peters file suit against Middlebury College for sexual discrimination after her being denied reappointment on the basis of her scholarship and teaching. The basic reason for the denial of reappointment was that she did not meet the standards of professional competence, especially in the upper division courses to which the English department aspired.

With regard to her (1) student evaluations, they were varied but with support for Peters predominating, but it was said that the students "ability to evaluate in the advanced and specialized areas of the English department was not established" (p.867). (2) The department chair sent a letter to the president of the college, saying, "The course of action I recommend is not likely to be popular with students who, though they in part recognize her intellectual limitation, are warmly responsive to her enthusiasm, energy, openness and ready human concern" (p.860). (3) The court ruled that "the reasons given by the institution for non promotion showed that by a preponderance of the evidence she was not professionally qualified for tenure and that the reasons given were not pretextual, and that there was substantial evidence to show that plaintiff's teaching was not satisfactory....(4) An evaluation of one's teaching ability is necessarily a matter of judgment. (5) A professor's value depends upon his creativity, his rapport with students and colleagues, his teaching ability, and numerous other intangible qualities which cannot be measured by objective standards" (p.860). Court found that she did not merit tenure, but was discriminated against.

IV. Lieberman v. Grant, 474 F.Supp. 848 (D.Conn. 1979).

Marcia R. Lieberman was hired by the University of Connecticut English department. Among her allegations was sex discrimination, but prior to her tenure review, she had been repeatedly warned by her department and its personnel committee that she had inadequacies in both teaching and scholarship.

(1) A five member departmental promotion and tenure committee all voted for tenure on the initial preliminary vote. After (2) being sent to a joint faculty committee on her department chair's negative recommendation (3) based on complaints received from "several students", to the effect that Lieberman's interest in feminism caused her to ignore other themes in literature, on close vote she was denied tenure; (4) a compilation of student ratings was showed that the cumulative ratings for members of the department ranged from a low of 4.09 to a high of 8.95. She had a cumulative rating of 7.06, which ranked her 12th out of the 15 junior faculty members. The 7.06 figure included the ratings from a previous semester in which the plaintiff received a rating of 8.18. Prior to this rating in the spring of 1972, the plaintiff's cumulative rating was 6.7; (5) Lieberman also attempted to introduce approximately ten personnel files concerning the tenure proceedings of other faculty in the English department for comparison; (6) Recognizing that such evidence would have had some minimal probative value, the Court, exercised its discretion under Fed. R.Ev. 403, and excluded it on the ground that "such probative value would be substantially outweighed by the delay and waste of time, which introduction of such evidence would have necessarily entailed....The plaintiffs case without such evidence seemed almost interminable, consuming 52 trial days over a two-year period. That is long enough" (p.873). The court concluded that sexual prejudice did not play any significant role in the tenure proceedings.

V. William Sypher v. Vermont State Colleges Faculty Federation (1982).

A faculty member contented that he was denied reappointment by Castleton State College, after his third year violated his academic freedom rights as stated in the faculty collective bargaining agreement. The faculty contended he was not reappointed because of his "political" activity on campus. The college contented he was not reappointed because of his inadequate teaching evaluations.

(1) Some of the student comments noted that, "When students try to disagree he shoots you down and trys to degrade you in front of the class," (p.115), while others said, "encourages student participation as much as possible... encourages student to express their ideas freely and not worrying how 'dumb' it may sound...always wants you point of view." (P.115) (2) With regard to the numerical ratings, the Board's opinion was that (3) regardless of a strong majority of student's rating his teaching as above average, (4) the existence of a significant minority of students feeling degraded, humiliated, and embarrassed can reasonably lead an evaluator to question a teacher's effectiveness; (5) the negative student evaluations are buttressed by the Chair of the Education Department; (6) were it not for this incident, it would be difficult for the Dean and President to support their subjective opinion that Sypher was not substantially above average in teaching effectiveness, given a statistical analysis of student evaluations demonstrating a strong majority of students believed to the contrary; (7) sufficient evidence exists from which the Dean and President could have reasonably concluded Sypher was not above average in his teaching effectiveness; (8) the Board went on to say that if they adopted the Colleges' view that Sypher was not reappointed because of his teaching effectiveness, no argument advanced by him

defending his teaching was likely to persuade the President because his decision was made on the "vigor and variety of student criticisms."

(9) With regard to the "political" aspect of the case, Sypher had written a letter in defense of his student rating level at the college which said, "it is certainly distressing when very good is not good enough, especially at a college with a modestly-talented student body that often discourages efforts at subtlety, wit and deeper penetration of subjects." (10) The Board responded to this letter saying, "other actions and statements by Grievant constituted legitimate reasons for not retaining him. In a May, 1980, letter to, Dean Beston, Grievant expressed his contempt for Castleton students" (p.135), (11) concluding, "Accordingly, we find credible the College's contention that Grievant was not reappointed because of his teaching effectiveness. [Italics added] (p.135).

VI. Lovelace v. Southeastern Massachusetts University, 793 F.2d 419 (1st Cir.1986).

Matthew Lovelace, a non tenured faculty at Southeastern Massachusetts University, challenged the decision not to renew his contract based on various constitutional claims, including a claim that the use of student ratings violated his academic freedom. He contended that his contract had not been renewed because he had refused to inflate his grades or lower his standards, and therefore his First Amendment academic freedom rights had been violated. More specifically, he maintained that in response to student complaints that his homework assignments were too time consuming and that his courses were too hard, the university first threatened not to renew plaintiff's contract unless he appeased the students. When he refused to lower his standards, his contract was not renewed.

In the court's ruling it said, (1) "It is important to note what plaintiffs first amendment claim is and to separate speech from action. Plaintiff has not contended that he was retaliated against simply because he advocated that the university elevate its standards.... Plaintiffs complaint instead is that he was retaliated against when he refused to change his standards" (p.425), (2) citing other cases, the court rejected his contention that university teacher has first amendment right to disregard established curriculum content, that the first amendment does not prevent a university from terminating an untenured faculty whose pedagogical style and philosophy does not conform to those of the school's administration, (3) "is a policy decision which, we think, universities must be allowed to set" (p.426). Further, the court ruled that (4) "We will assume for purposes of this opinion that plaintiffs refusal to lower his standards was a substantial motivating factor, (see Mount Health Board of Education v. Doyle, 429 U.S. 274, 283-284, 97 S.Ct. 568, 574-575, 50 LEd.2d 471 (1977), in the decision not to renew his contract...Whether a school sets itself up to attract and serve only the best and the brightest students or whether it instead gears its standard to a broader, more average population is a policy decision which, we think universities must be allowed to set....matters such as course content, homework load, and grading policy are core university concerns, integral to implementation of this Policy decision. (p.424).

VII. Fields V. Clark University (1987).

Fields (Neither first name nor academic degree level was cited) taught at Clark University in the sociology department. While Field's scholarship and contributions to the university were considered adequate, they were deemed by various reviewing groups not to be so outstanding as to overcome deficiencies in her teaching. Fields contended she was being discriminated against on the basis of her sex.

In support of the collective opinion that (1) Fields was a poor teacher, the university (2) submitted a group of student evaluations, (3) a few of which, from students in Fields' seminars, were "wildly enthusiastic" about her enthusiasm, commitment and presentations; (4) a few were ambivalent; (5) with a considerable number being extremely negative, particularly (6) with regard to her large lecture classes in basic courses in sociology. The students (7) complained that the plaintiff's courses were disorganized. that there was no correlation between reading assignments and the course work and (8) that the plaintiff herself was often unprepared and (8) unresponsive to students' needs, with (9) one student saying that her performance was so bad that it was a principal reason for her leaving the university.

The court noted that (10) Fields' "attacks" the university's use of her student evaluations because they were not gathered and evaluated according to accepted standards of scientific polling procedures. In response, the court agreed, saying, "She is probably correct. The use made of the student evaluations in her case, however, followed the practice at the defendant university in other tenure decisions" (P.671). On the findings of sexual discrimination the court ruled in Fields' favor.

VIII. Carley v. Arizona Board of Regents, (1987).

Denny Carley was denied reappointment by Northern Arizona State University where he taught in the department of art as assistant professor. The denial was largely on the basis of student evaluations of his teaching. This appeal is from a superior court judgment affirming a decision of Northern Arizona University President to deny the renewal of a teaching contract to Carley. The issues on appeal involved: (1) whether his right to academic freedom was violated because student evaluations were utilized as the primary tool to determine his teaching effectiveness, (2) whether the President abused his discretion by rejecting the findings of the majority of the Academic Freedom and Tenure Committee and (3) whether there was substantial evidence to support the President's decision.

In term of his teaching, (1) of the 13 faculty in his department of art, he was ranked fifth, (2) by his chairman he was ranked seventh, (3) student evaluations, however, ranked him last: 13th of 13. (4) Carley identifies as "protected speech" his teaching methods where his goal in his commercial art course was to promote a business atmosphere by requiring attendance, promptness, and self-reliance, and required them to meet deadlines. (6) He considered this approach part of the content of his class. (7) He characterized his professional style as being a "demanding teacher contrary to some student expectations." (8) Because of this, he maintained his popularity suffered as reflected in his low student evaluations. (9) An examination of student comments on the evaluation forms indicated that Carley was correct in his assessment as 61% (49 out of 80) negative student comments focused on these values.

(10) Carley maintained that because student evaluations were critical of those methods, the student evaluations challenged his academic freedom rights, concluding that the student evaluations cannot be used as the primary basis for failing to renew his contract (p.1101). (11) With the ACLU, Carley filed suit, with his attorney saying, "Our primary argument is that it's a violation of academic freedom---he wasn't doing anything that wasn't in line with being a demanding

teacher.

(12) A departmental review committee voted for non-reappointment, (13) but the department chair disagreed, recommending to the dean that the appointment be renewed. (14) The University-wide Academic Freedom and Tenure committee, with a 6 to 3 vote, also decided in the Carley's favor, finding "improper use of the student evaluation instrument, a de facto violation of academic freedom, [and] personal prejudice" which "constitutes substantive violation of due process." (15) They further found that the judgment of his teaching effectiveness had been "based wholly on student evaluations" which have "no objective validity as a measure of teaching effectiveness." (See Heller, 1986, italics added) (16) Three dissenting members of the committee filed a minority report, finding that his department colleagues had treated the faculty member fairly and that student evaluations were only one of several criteria in judging his teaching effectiveness.

(17) The president reviewed Carley's case and found justification for not supporting his reappointment. The court ruled (18) the University president was free to consider factual findings made by minority members of academic freedom and tenure committee and any other evidence which he found relevant in determining whether to deny renewal of teaching contract to non tenured instructor. The president was not bound by factual findings made by majority members of committee.

(19) It was said that Carley "paints with too broad a brush the concept of academic freedom," since his teaching style is not a form of speech protected under the First Amendment. (20) Decision not to retain a non tenured instructor, even if based, in part, upon student evaluations expressing disapproval of his teaching methods, did not violate instructor's First Amendment right to academic freedom; (21) Carley was not denied a contract because of expressing unpopular opinions or otherwise presenting controversial ideas to his students. (22) Thus, we conclude that the decision not to retain Carley, even if based, in part, upon student evaluations expressing disapproval of his teaching methods, did not violate his first amendment rights. p.1103. (23) Carley has cited no authority that relying primarily or solely on student evaluations would be impermissible. We have found none. p.1105 Denial of reappointment was upheld.

IX. Robert Kramer v. The President of the University of British Columbia (1992).

Kramer, an American citizen, was Assistant Professor in the Department of Asian Studies, Faculty of Arts, whose first two-year reappointment was denied. Kramer maintained that there were errors of procedure, failures to obtain relevant evidence, and that the denial was based on a defective dossier.

(1) The department vote was 4:3, against reappointment. The department Head also did not recommend reappointment. (2). The department Head viewed Kramer's 1989-90 course evaluations "with some alarm." (3) The numerical scores were among the lowest of the tenure track faculty. (4) Even more disturbing to the department Head was that a considerable number of students in their written comments stated that Dr. Kramer was biased sarcastic, and hostile to the material and that a number of students had stated that Dr. Kramer's teaching would cause them to stay away from the Asian Studies department. (5) There were also some diametrically apposed positive comments. (6) Since the course was the general introduction to the subject and the Department, such negative comments were of great concern to the department Head.

(7) Kramer argued that the most significant mistake was the failure to consider all aspects of his teaching, as outlined in Article 4.02. (8) Only his teaching in 1989-90 was considered, whereas (9) he had taught a wide range of courses over the previous three years (10) had three new courses that year, (11) plus a graduate course. The basic course, Asian 105 had 220 students, making it numerically the largest class per faculty member in the department (Even though it was shared between two instructors). (12) his graduate teaching was entirely ignored, and (13) only the formal student evaluations were considered, with no peer evaluation. (14) Student evaluations were considered from the standpoint of his popularity, not his effectiveness. (15) Only one of the over thirty numerically rated questions was use: "Rate instructor bad to good". (16) While a number of negative student comments were quoted in the department Head's letter, there were a number of very positive comments, and these were not mentioned at all. (17) Department head indicated that his teaching was not up to the departmental "standard". The standard appeared to be the performance of the tenure-track faculty, though Kramer was one of the most junior faculty members.

The board of appeals responses included, (18) "The most important perceived error in the teaching evaluation, in the opinion of the Board, is the reliance solely upon the student evaluations and written comments for the 1989 course evaluations. There was no peer review at all; no member of the Department audited any of Dr. Kramer's lectures. There was, therefore, nothing to guide the Department but the student comments," and "no way to test the accuracy or fairness of the undoubtedly disturbing comments in Asian Studies" (p.10). (19) "Given that the Departmental procedure is to set aside a class hour for class discussion without faculty present, with the written evaluations and comments made during the same hour, there is a danger that some negative class commentary will dominate the discussion and will not be the 'independent' opinion of all of the students. (20) This is especially true in the context of the direction to assess "effectiveness" versus "popularity" (p.10). However, the board noted (21) "As for the 'popularity vs. effectiveness' debate, a discouraging or hostile attitude is a part of effectiveness as much as it is of popularity" (p.8). (22) "We recognize that the strong negative comments...may have [been] received...too late in the first year to have any peer teaching assessments in that particular class, but there does not appear to have been any attempt to see if the apparent problem had continued in September or October of 1990 in any courses, or (23) if the student perception was accurate and fair" (p.10).

The board further noted, "In the result, one got a 2.82 and one got a 3.07, and Dr. Overmyer cautiously and generously assigned the higher rating to Dr. Kramer the difference is statistically invalid in any event. But any rating at all in this course is suspect. (24) Further, it is the largest class in the Department, and large classes generally attract lower scores. Dr. Kramer's scores in the other two courses were higher--3.45 in one, 3.91 in another, against a "faculty average" of 4.22. As his counsel pointed out, that is a comparison against regular faculty, mostly senior and tenured" (p.10-11). (25) "We have examined all of these written comments. There was a very wide range of comments. There were not 29 comments saying sarcastic and biased; but there were certainly 29 comments which included either cynical, sarcastic, biased, insulting, negative, condescending, belittling, opinionated, arrogant, nihilist, and destructive... [but] (26) It was obvious that almost all of the class were upset about an examination which was considered more geography than Asian Studies, and (27) they didn't like the marking. (28) They also felt the workload was far too heavy for an "introductory" course. (29) However, it would only be fair to add that there were a number of comments in favour of Dr. Kramer, stating that the student "liked the course immensely", "now interested in Asian Studies"; "helps create a relaxed atmosphere", "really enjoyed him", "very approachable and knowledgeable", "very enthusiastic", "captivates audience with his humour", "very effective" (p.12). (30) "In the other two courses, both small, both Japanese language, there were also some negative comments. One perceptive student noted that some of the unhappiness came from the fact that the levels of Japanese language ability were badly divided; some found it easy, others very hard" (p.12).

The board concluded, "In the final analysis, we feel that this review of the Head's comments on teaching, which would be the sole evidence upon which the Dean and the President could rely, shows that it was incomplete and might have been misleading, due to the total focus on student evaluations and the lack of peer opinion on teaching ability....The original Departmental vote was so close, we may find it easier to conclude that a different result might (We do not say must) have occurred, and that we should return the case: here. to the departmental level" (p.12-14).

X. Christopher Turner v. The President of the University of British Columbia (1993).

Christopher Turner, who taught Russian literature was denied promotion to full professor. In not recommending promotion, the Departmental review document in part stated:

(1) Turner's main issue was that there were a number of inaccuracies and misleading descriptions in the critiques of his teaching, with teaching weaknesses the main reason for the negative decision on his promotion, noting (2) "the small number of student evaluations handed in from his undergraduate literature course (3, 8, and 6). Turner provided the Board with the class lists in those courses, showing that there were 4, 16, and 16 students who had registered in those courses, and that 4, 14 and 16 had completed them. Thus the actual numbers were proportionately much higher than the 'responses' referred to... (3) It was argued that a low 'response' rate was proof that rather few students were actually attending, at least on the date of the evaluation, but that is a dangerous inference; some students simply do not fill in responses; some students in courses with no final exam may not attend the final evaluation lecture. (4) It would be easy for readers to infer from the [Department Head's] figures that Dr. Turner's classes were smaller than comparable classes in the department, a conclusion which Dr. Turner was able to rebut with course enrolment figures. In fact, it is clear that [the dean] misread [Department Head's] figures; in her letter of reasons to Dr. Turner...she states 'there were few students in undergraduate literature courses since 1986/7---(3,8, and 6 respectively...', and thus has taken 'response' figures for actual enrolment. Thus the general criticism in the Head's letter and the Dean's later letter of reasons, about relatively small classes compared with others in the department, was not supported in the material available to the University. (5) This misunderstanding is in our opinion sufficient in itself for a reconsideration, since teaching was the focus..." (p.3).

The board noted that (6) "While there is no question of Dr. Turner's competence as a teacher at all levels, teaching evaluations for the last several years show that his effectiveness is marred by what students perceive as excessive formality, lack of enthusiasm and dullness....In a previous promotion attempt, his teaching was briefly described as "very competent" but student evaluations indicate further improvement to be "better than adequate." (p.2). (7) "while not ignoring some student unhappiness with Dr. Turner's teaching style, we think that the comments and emphasis on the size of Dr. Turner's classes as evidence of poor teaching are open to objection and constitute errors of procedure and/or evidence" (p.6).

The board further noted that (8) "This board has been asked on a number of occasions to pass judgment on the relevance of student evaluations to the Agreement criteria for good teaching. Good teaching is an elusive concept. Students may not be good judges during a course; their judgment might be quite different several years later in life. On the other hand, while popularity is not competence nor effectiveness, to the extent that it encourages students it has some relation to both" (p.7). (9) While the Agreement permits, but does not mandate either student reviews or peer reviews, and the methods of assessment 'may vary', we do conclude that the reliance placed on these very limited student reviews must have been great, since there was no other evaluation referred to. Where there is no other evidence sought, student comments will have an apparent importance and credibility that they may not deserve... (10) We would strongly recommend peer review in the reconsideration which we are requiring" (p.7).

The board concluded that (11) "there were sufficient errors of procedure and/or evidence to return the case for reconsideration" (p.11).

XI. University of Regina Faculty Association v. University of Regina (1993).

Dr. Pradeep Jalan was an Associate professor in the Faculty of Administration at the University of Regina. In February 1993, during Jalan's fifth year at the University, the Faculty Review Committee recommended that he be granted tenure. This recommendation was not accepted by the Dean, who recommended denial of tenure on the basis that Jalan had not met the standard for an Associate Professor in the areas of scholarly performance and teaching.(Education Employment Law News, 1994). Jalan appealed the Dean's recommendation to an appeals committee, which denied the appeal. The University of Regina Faculty Association then filed a grievance on Jalan's behalf, claiming he had been wrongfully denied tenure. The University justified its denial of tenure on two grounds. (1) that Jalan's teaching was unsatisfactory, submitting student evaluations completed over the course of five years in support of this conclusion, and (2) arguing that he had not met the standard of scholarship required for tenure, as Jalan had only two publications, neither of which had been published in a refereed journal; they appeared only in publications of conference proceedings.

The Board ruled, (3) "With respect to teaching, it is our opinion that the evidence of unsatisfactory performance is very weak indeed ...It is important to note that the basis of the comments, particularly the negative ones in the fail of 1992, were written student assessments...[and] Although these assessments are expressly recognized in Art. 17.19 of the collective agreement, to base important career decisions on them only does not seem justified" (p.4). The Board further ruled (4) that tenure decisions could not be based solely on assessments which were completed by students who had never been made aware of the ramifications of their statements. (5) [I]f evaluations are to be used for serious career development purposes those completing them should be aware of the potential consequences of their participation" (p.4) The Board argued (6) that the University was under an obligation to verify negative comments before acting on them. Consequently, (7) the fact that Dr Jalan had received some negative evaluations from students could not be used to undermine the otherwise generally favorable comments he had received in his annual performance reviews. (8) "To base serious career decisions narrowly on student evaluations is not to be encourage...Indeed, in a document prepared by the University... titled, 'The Assessment of Teaching Effectiveness: Guidelines for Teaching Evaluations' it is recognized that teaching evaluations can involve a variety of techniques. The document states that: 'These may include direct observations, informal feedback from students (both past and present), student course/instructor evaluations, review of instructional material selected by the instructor, assessment of the appropriateness of textbooks, examination of class outlines, assessment of the appropriateness of tests and examinations set by an instructor, and grading practices.' In the case before us, only one of these techniques was relied upon. If (9) teaching is to be seriously evaluated for career purposes, whether for positive or negative purposes, it seems incumbent upon Faculties not to rely only on classroom administered evaluations but to broaden the base of assessment" (p.4) The Board said teaching was wrongfully evaluated, but upheld denial of tenure on grounds of inadequate scholarship. University of Regina V. University of Regina Faculty Association And Dr. Pradeep Jalan (1993).

With respect to teaching, it is our opinion that the evidence of unsatisfactory performance is very weak indeed. It will be recalled that in the first two years of his appointment there were negative comments with respect to Dr. Jalan's teaching. However, in year three and four comments were quite positive. Then in the fall of 1992 again the comments were negative. It is important to note that the basis of the comments, particularly the negative ones in the fall of 1992, were written student assessments that were universally applied in the Faculty of Adhinistration. Although these assessments are expressly recognized in art. 17.19 of the collective agreement, to base important career decisions on then only aces not seem justified. Indeed, in a document prepared by the University dated February 15, 1991 titled, "The Assessment of Teaching Effectiveness: Guidelines for teaching Evaluations" it is recognized that teaching evaluations can involve a variety of techniques. The document states that:

> These may include direct observations, m formal feedback from students (both past and present), student course/instructor evaluations, review of instructional material selected by the instructor, assessment of the appropriateness of textbooks, examination of class outlines, assessment of the appropriateness of tests And examinations set by an instructor, and grading practices (p.19).

20. In the case before us, only one of these techniques was relied upon. If teaching is to be seriously evaluated for career purposes, whether for positive or negative purposes, it seems incumbent upon Faculties not to rely only on classroom administered evaluations but to broaden the base of assessment. The University itself has recognized this in the February 15th, 1991 document. To base a serious career decision narrowly on student evaluations is not be encouraged. That is not to say that student evaluations properly structured and properly administered cannot be an important part of the evaluation but, as the University has recognized, it should not be the only factor. This R is particularly so when, as was strenuously argued by Ms. Rasmussen for the Association, the students are not advised of the potential use C! the evaluation tool. The point which is Very valid, is that if the evaluations are to be used for serious career development purposes these completing them should be aware of the potential consequences of their participation.

It is our conclusion that the grievor met the onus placed on him to establish oh that he was a competent teacher. This onus was met through the statements made in the Annual Performance Reviews and the evidence relied upon by the employer to suggest that the standard was not met, in our view, was not sufficient.

XII. Cynthia J. Fisher v. Vassar College, (1995).

Dr. Cynthia Fisher taught in the biology department at Vassar College. An appeal by defendant Vassar College from final judgments of the United States District Court for the Southern District of New York (Motley, J.). Following a bench trial, the district court found that, in denying plaintiff tenure, defendant had discriminated against her by reason of her sex and age.

An analysis of Fisher's teaching ability included a review of her student evaluations, which (according to the report) (1) reflected "consistent problems with clarity and her ability to illuminate difficult material" but were otherwise generally positive; the district court (2) found that the biology department distorted Fisher's teaching recommendations by (3) "selectively exclud[ing] favorable ratings", by "focus[ing] on the two courses in which Dr. Fisher had difficulties" and (4) by "applying different standards to her than were applied to other tenure candidates" (p.1209). Further, the district court observed (5) that "the males tenured while Dr. Fisher was on the faculty were praised for their teaching while Dr. Fisher was criticized for her teaching, despite (6) the facts on which the Committee's determinations were based (student evaluations, Biology Majors Reports and [Student Advisory Committee] reports) revealing that Fisher's evaluations were superior to theirs" (p.1211); the court noted that (7) statistical analyses may be a part of a plaintiff's effort to establish discriminatory treatment. Finally, (8) the court . . . found that the termination of Fisher's employment resulted from a pretextual and bad faith evaluation by Vassar of her qualifications, not from any inadequacy of her performance, qualifications, or service.

XIII. Yu Chuen Wei v. Vermont State Colleges Faculty Federation (1995).

Dr. Yu Chuen Wei taught in the Mathematics department at Castleton State College where she was denied tenure. The case was arbitrated by the Vermont Labor Relations Board (VLRB 261, 1995), Wei's case involved SEF as a part of a wider issue of racial discrimination. An extensive review Wei's student evaluations were described by the Board.

With regard to her student evaluations, (1) Wei contended that the College acted unreasonably and arbitrarily by denying her tenure on grounds of teaching effectiveness without first performing a systematic statistical comparison of the student evaluations. (2) She brought in an expert who had researched and written an article on attitudes of student as a source of contamination on SEF to testify that there were serious problems with the validity of SEF under the best of statistical circumstances. (3) As an expert in the field of statistical assessment of SEF, her findings indicated that SEF relative to the quality of instruction is significantly affected by inappropriate attitudes, thereby creating suspicion about the objectivity of undergraduate student evaluators. (4) In specific, the expert found: (a) the race, sex or ethnic background of the professor; (b) the grade the student expected to receive; (c) whether the course was a required one; and (d) whether the student an undergraduate or graduate student influenced how students evaluated professors. (5) In analyzing Wei's evaluations, the expert compared the mean or average responses of three years of SEF questions for Wei to the mean or average responses to the same questions for a male colleague for the same period of time who was the most recently tenured professor in the Mathematics Department. Both an older and a newer evaluation form were analyzed. (6) A statistical comparison utilized a t-test. A t-test is standard for determining whether the difference in two means is significant, i.e., to determine whether differences are valid or due to chance or random error. (7) The results indicated: For (8) upper level classes using the "old" evaluation form, Wei's mean responses were significantly higher than the male colleague's mean responses in four out of the five questions pertaining to instructor attributes. (9) In one category, there was no significant difference between Wei's mean value of responses and the male colleague's mean value of responses. For (10) lower level classes using the "old" evaluation form, in two categories there was no significant difference between Wei's mean value of responses and the male colleague's mean value of responses.

(11) In three categories, the male colleague's mean value of responses was higher than Wei's. In (12) upper level classes using the "new" evaluation form, Grievant's mean value of responses was significantly higher than that of the male colleague for three of the questions. (13) In the other two categories, there was no significant difference between Wei's mean value of responses and the male colleague's mean value of responses. For (14) lower level courses using the "new" evaluation form, the male colleague's mean value of responses was significantly higher than Wei's in two categories. (15) In the other three categories, there were no significant difference between Wei's mean value of responses and her male colleague's mean value of responses. (16) Wei maintained that her students rated her the same or higher than the male colleague's students rated him.

The Board disagreed (17) "that the statistical comparison of Grievant's student evaluations with those of [male colleague] contributes to a conclusion that the Colleges' articulated reasons for denial of tenure and promotion constituted a pretext for discrimination. (18) We note that the comparison offered by Grievant is somewhat weak since [male colleague] was tenured in 1988, and those student evaluations of his which were compared with Grievant post-dated his tenure review by a number of years."

Moreover, they said, (19) "The statistical comparison demonstrates that Grievant was evaluated higher by students than [male colleague] with respect to upper level classes, but that (20) [male colleague] was evaluated higher than Grievant in lower level classes. Given (21) this "mixed" result, the statistical comparison of evaluations does not demonstrate by a preponderance of the evidence that Grievant's students rated her the same, or better, than [male colleague]" (p.305). Further, the Board said, (22) "the statistical comparison is insufficient by itself to demonstrate how students evaluated the respective faculty members. The statistical comparison does not take account of the comments made by students on the evaluation forms. Grievant's student evaluations are striking in how often mention is made of Grievant's communication difficulties, particularly language difficulties. No similar problem, or as serious a problem, appears on [male colleague's] student evaluations with the frequency in which Grievant's communication difficulties are mentioned....In any event, (23) even assuming the statistical comparison was valid, the comparison does not support the conclusion that Grievant was treated in a discriminatory manner compared to her male colleagues." (p.304-5). The board judged that (24) the tenure criteria "are not drawn with mathematical nicety." The board further ruled that (25) the Dean and the President, both reviewed Grievant's student evaluations carefully. Their failure to take it a step further, and perform a statistical comparison of Grievant's student evaluations with those of other faculty members who have been granted tenure was not arbitrary and was reasonable; (26) Such a comparison is nowhere required by the Contract, and (27) we decline to hold such an involved comparison is necessary before a reasonable tenure determination can be made; (28) The Dean and the President obviously had much experience in reviewing student evaluations, and could reasonably draw on that experience in each tenure review. (p.311).

The Board went on to say, (29) In many other evaluations, in which students refer to communications problems in Grievant's teaching, references to difficulty in understanding Grievant may reasonably be interpreted as expressing a concern about her ability to communicate" (p.305). Finally the board noted that while some students had written that she was a "slant eyed bitch," and the she should "go back to China," (30) "We also are not persuaded that the racism evident in the student evaluations of Grievant made student evaluation results unreliable. The percentage of evaluations in which racism by students was evident was approximately one percent of the total evaluations." (p.306)

(31) Wei's last claim charged that the College violated the Contract by denying her a promotion, even though both her scholarly performance and professional activities were exceptional. Article 22(E) of the College provides for otherwise granting promotion if the President decides that "performance in one of three areas has been exceptional" (p.314).

With regard to this Article, the Board concluded that (32) "Grievant has established no such discrimination. Although Grievant had a significant publication record, most of it was developed before coming to Castleton" (p.315). (33) In terms of exceptional scholarship, Dr. Wei maintained she had solved a significant mathematical problem (apparently published). The Board's response was, (34) "although Grievant claimed to have solved the Erdos conjecture, Dean Mark reasonably concluded that she had not established that she actually had solved the conjecture. Under these circumstances, and given our consideration of the discrimination issue previously discussed, we conclude that (35) Grievant has not established discrimination. The Colleges reasonably, and based on legitimate reasons, concluded that Grievant had met the tenure standards in this performance area but that her performance was not exceptional" (p.315). (36) The Board concluded that Wei had not established that the tenure decision in her case was unreasonable, arbitrary or based on erroneous reasons.

XIV. Guam Federation of Teachers v. The University of Guam (1990).

At the University of Guam, a ruling to remove anonymous student evaluations from professors' tenure files was handed down by an arbitration board as the result of a rare challenge to the use of such evaluations in tenure and promotion decisions (Blum, 1990). The action was in response to a grievance filed by the university's faculty union, the Guam Federation of Teachers, which is affiliated with the American Federation of Teachers (AFL-CIO). The union said the use of SEF violated the union's contract with the university, which provides that anonymous documents or those "based on hearsay" should not be included in a faculty member's file. The union also argued that the university improperly interpreted the data from the evaluations.

Some of the issues here are (1) students not being made aware of the purpose and ramifications of their evaluations, (2) the anonymous nature of student evaluations, (3) the invalid analysis of SEF, and therefore, (4) SEF in effect being anecdotal and hearsay data. Since most SEF results are prepared anonymously, an instructor has no recourse to confront his/her evaluators.

XV. Dr. Brian Maclean v. President of The University of British Columbia (1991).

Held at Vancouver, B.C. January 28--June 20.

Dr. Brian MacLean was an Assistant Professor in the Department of Anthropology and Sociology, whose first two-year term of employment was not renewed.

(1) The Departmental Committee voted 16:1 against renewal, as did the Head, the Dean's Advisory Committee, the Dean and the President. (2) With three years of teaching experience in Saskatchewan, and London, his scores for Anth/Soc. 100 on one of the critical questions--- "Would you take another course from this instructor" were very low--29%(fall) and 34% (spring), compared with an average of 69% for all other instructors in the course. (3) They were the lowest for all other sections of the course, and for all first and second year courses. (4) Moreover, he had the lowest scores on another important question--- "would you recommend this course to a friend". (5) In his speciality course, Introduction to Criminology, his scores on the above questions were 61.2% for the instructor, 74.6% for the course, which, according to the Department Chair were well below other specialized courses. (6) In his senior course, in Criminology, his ratings were 60% for the instructor and 40% for the course [precise rankings were not provided] (7). When MacLean was not awarded a "merit" salary increase, the Chair told him that this was because of his low teaching ratings.

(8) His Department Chair testified to the Appeals Board that this problem was discussed with Dr. MacLean and it was strongly suggested he consult with "good" teachers in the department. (9) The Chair stated that MacLean's response to the bad ratings was that he did not believe them, that (10) he denied that there was a problem, arguing that "the students were not very good students" (p.6).

(11) Peer reviews of his teaching were conducted. (12) Before the classroom visits, the two reviewing faculty members had been informed by the Dean that Dr. MacLean had a "teaching problem" the previous year. (13) Both faculty then attended a large lecture about 80 students. (14) Both reviewing faculty reported a favourable opinion of the lecture, noting that Dr. MacLean's speech was clear, his "delivery" was good, (15) despite the disadvantage that it took place in one of the more "difficult" and unpopular lecture rooms. (16) Neither faculty attended lectures in Dr. MacLean's other two small courses because they had only four students in each and (17) "hence were not amendable to the usual 'lecture' style and could be easily disrupted by a visitor" (p.7). (18) Instead, in these two courses, a class visit was made in the absence of Dr. MacLean.

(19) In addition to the student questionnaires, the reviewing faculty held in-class discussions about his teaching. (20) In general, the in-class peer reports were mixed but favourable. The in-class discussions were more problematic. (p.30). (21) While the knowledge, interest and enthusiasm of Dr. MacLean was acknowledged, "the problem appeared to be one of style or personality." (22) He was criticized for "putting them down", for belittling their responses to his questions in class, and that he was "hostile, aggressive, volatile, biased" (p.9). (23) It was noted that, "The students had obviously been talking among themselves before this class visit". (24) One of the reviewing faculty asked the Dean if the Departmental meeting could be postponed until after grades were submitted so that the students would feel protected. Inquiries to the President's

office found that this was acceptable. (p.9).

The court noted that (25) "With respect to the "qualitative" scores---i.e. the "comments", there was a clear error. The qualitative comments from a number of courses were read and commented on, and conclusions were drawn from them which went into the "file". Both Reviewing faculty read and commented on them, as did the Department Chair in her letter to the Dean. Yet the Dean had clearly stated in a departmental memo that the qualitative comments were not to be used for administrative or promotion purposes. In evidence she admitted that they were used, and indeed as we have said they formed parts of the "file". (26) While in the abstract there is no reason why such comments would not be relevant, if the Department had a rule against their use, or in other words if they were "for the professor's eyes only", then it was a significant breach of Departmental rules to use them" (p.31). (27) In the opinion of the Board, so long as the comments were fairly presented, they offered the PAT [Promotion and Tenure Committee] and others a better balanced view of the teaching qualities and problems of Dr. MacLean than the quantitative statements alone" (p.31). (28) The court noted that "One problem with the questionnaire is that it solicits bad points as well as good points. Despite that caveat, we conclude that the inclusion of the qualitative comments was not a significant error." (p.32).

It was further noted that (29) "As against the low figures, they disclosed a number of good qualities in Dr. MacLean---enthusiasm for his subject, wide knowledge of the literature, much out of class assistance to students, and a commitment to seeking good work from students.(p.31). (30) The reviewing faculty report noted the comments about Dr. MacLean's "derogatory manner, biased opinion, unwillingness to listen", were matched by "clear, stimulating, very helpful after class". And, (31) "some students have told us that the comments made were not representative of the class as a whole and were unduly influenced by the process" (p.41). (32) "A number of students, both from earlier years and from his current classes, furnished letters of support, and in preparation for the appeal, some furnished affidavits with respect to particular matters such as the 'intimidation' discussion in Soc. 250 and events in Soc. 490 and 520 in the fall of 1989." (p.33). (34) "We agree with the appellant that these comments are of some importance, since we agree that while a superior research and publication record cannot overcome a poor teaching record, it might tip the scales where the teaching record was 'on the edge'" (p.10).

(35) The Faculty Agreement specified that "Evaluation of teaching shall be based on the effectiveness rather than the popularity of the instructor, familiarity with recent developments in the field, preparedness, presentation, accessibility to students and influence on the intellectual and scholarly development of students" (p.30), (36) MacLean sought expert opinion about the validity of the student evaluation instrument. Both MacLean and the University called expert witnesses to testify as to the reliability and relevance of the evaluation instrument. (37) The evidence, submitted by two experts, one for MacLean and one for the university, was conflicting. (p.30).

With regard to SEF validity, the court concluded (38) "that the instrument was not perfect, that it had flaws, and that the very limited number of samples (because of the very limited number of courses and students surveyed over the period) impaired its reliability. (p.30). (39) "However, we accept the evidence of Dr. [X] that the instrument has some value, directed toward the specified factors. (40) The relevance and quality of the scores is a matter of weight for the various decision-makers, and we assume that they were reasonably aware of the limitations of student evaluations and gave them the weight they deserve" (p.30).

Finally, the court said, (41) "As we have said, there were some errors of procedure and of evidence. We have set them out in some detail; and although they have some substance, they must be viewed against the much larger body of evidence which was either not questioned or was confirmed during the hearing" (p.38). "The evidence before us led us to conclude that Dr. MacLean has worked hard, has been active in research, and has some teaching talents of value. (42) It is clear that he appeals to a proportion of students who find him stimulating and helpful. (43) A number were willing to testify on his behalf or to supply written testimonials" (p.38). (44) "Dr. MacLean's faults appeared to be faults of style or approach...(45) Nonetheless, his colleagues have made a judgment based upon what we have found to be an adequate foundation of evidence" (p.38).

The court upheld the university.

About the Author

Robert E. Haskell

Brief Bio ~ Robert E. Haskell has been teaching college and university level courses for over twenty years. He earned his Ph.D. from the Pennsylvania State University in Psychology and Social Relations, his M.A., and B.A. from San Francisco State University. His areas of research and teaching include: transfer of learning, analogical reasoning, small group dynamics. Major publications include: four books, the latest of which is, *The Future of Education and Transfer of Learning: A Cognitive Theory of Learning and Instruction For The 21st Century* (forthcoming), and numerous presentations, chapters, and research articles in national and international journals. He also serves on several editorial review boards, and is Associate Editor of The Journal of Mind and Behavior. He is former Chair, and currently Professor of Psychology, Department of Social and Behavioral Sciences, University of New England.

Professor of Psychology
University of New England
Biddeford, Maine 04005

UNE Home Page: http://www.une.edu/sbs/haskell.html

E-mail: rhaskell@javanet.com

Copyright 1997 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://olam.ed.asu.edu/epaa

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, Glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692)

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Marshall University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Rocky Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
Arizona State University

Dennis Sayers
University of California at Davis

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina--Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher Education

Mary P. McKeown
Arizona Board of Regents

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois--UC

Robert T. Stout
Arizona State University

archives | abstracts | editors | board | submit | comment | subscribe | search

This article has been retrieved 1377 times since August 25, 1997.

Education Policy Analysis Archives

Volume 5 Number 18 August 25, 1997 ISSN 1068-2341

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass Glass@ASU.EDU.

College of Education

Arizona State University,Tempe AZ 85287-2411

Copyright 1997, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

Academic Freedom, Promotion, Reappointment, Tenure And The Administrative Use of Student Evaluation of Faculty (SEF):

(Part III)

Analysis And Implications of Views From The Court in Relation to Accuracy and Psychometric Validity

Robert E. Haskell [1]
University of New England

Abstract: In two previous papers, it was noted that while a controversial history of research on the reliability and validity of student evaluation of faculty (SEF) exists, it has not been typically viewed as an infringement on academic freedom, promotion, reappointment, and tenure rights. As a consequence, legal aspects of SEF are neither readily apparent, nor available. Legal rulings, their implications and assumptions in relation to their accuracy and psychometric validity where SEF are integral to the denial of academic freedom, tenure, promotion, and reappointment are reviewed along with the legal principles of Disparate Treatment and Disparate Impact, and the scientific Precautionary Principle in policy decisions.

Table of Contents

- Brief Overview of the Validity of
- Procedural, Burden of Proof, and

SEF
- **The Courts' Approach to the General Accuracy and Psychometric Validity of SEF**
    - Historical Overview of the Courts' Approach to the Validity of Faculty Evaluation Data
    - Acceptance of Administrative Subjective and Untrained Evaluator Judgements Of SEF Data
    - SEF as Social Judgement and Diagnosis
- **Variables Affecting Validity Not Taken Into Account When Assessing SEF**
    - Instructional Variables
    - Student Biases Variables
    - Popularity Variables and Effectiveness
- **The Courts' Reliance on Both Quantitative Data and Qualitative Comments in SEF**
    - Reliance on SEF v. Peer Evaluation
    - Numerical Ranking of Faculty
    - Use of Qualitative Written Student Comments
    - Mixed Student Comments
    - Transcendent Value of a Professor Over Teaching Quality

Policy-Decision Criteria in Assessment of SEF
    - Validity Assessment of SEF as Procedural or Process Issue
    - Decision Criteria and the Scientific Precautionary Principle
- **The Court's Approach to Validity of SEF in Relation to the Principles of Disparate Treatment and Disparate Impact**
    - Disparate Impact
    - The Disparate Treatment and Impact Principles Generalized
- **Beyond Statistical Significance of SEF Research**
    - Assumption # 1: Statistical Significance of Indicators of Teaching Effectiveness
    - Assumption # 2: Statistical Significance of SEF of Teaching Effectiveness Measures
    - Appropriate Learning
- **Conclusion**
- **References**
- **Appendix: A Non Litigated Case of SEF Used in the Denial of Tenure and Reappointment**

As I indicated in previous papers on SEF (Haskell, 1997a, 1997b), the history of legal rights demonstrates that issues not considered to have legal standing only come to have legal standing after a long process of advocacy. The evolution of a policy or legal principle requires the accumulation of data, coalescing judgements and arguments. To this end, this paper, will continue to examine court reasoning and rulings on SEF in cases involving the denial of academic freedom, tenure, promotion, and reappointment (AFTPR) decisions in relation to its implications and assumptions regarding accuracy and psychometric validity.[2]

In a second paper on SEF, (Haskell, 1997b), I abstracted from the text of located legal cases views from the court pertinent to SEF. The appendix of the second paper provided a verbatim abstracting of the text of each case relative to its SEF content. As a consequence, in summarizing the pertinent findings of that paper for the present one, for convenient referencing the specific case textual material for each section will be in placed in a footnote indicated at the beginning of each section heading and the indented "Summaries" are carried over from Part II.

For convenience, I will use these abstracted legal views and rulings to examine their implications for the courts use of SEF in relation to their accuracy and psychometric validity. A final paper will address the implications of court rulings for academic freedom and instruction. As I noted in my second paper, not only are legal cases prima facially complex, but when specific legal definitions (e.g., disparate treatment and impact) and other special Congressional Acts (e.g., EEOC) are superimposed on them, they become logically unwieldy, not just to the non legal scholar, but apparently to the Courts as well.[3]

Finally, I would like to point out that the issues examined in this series of papers are not primarily concerned with individual faculty rights but with the implications of SEF when used for administrative purposes on academic freedom, educational quality, standards, and ultimately on the competence of graduates.[4]

Brief Overview of the Validity of SEF

As shown in Haskell (1977b) and reiterated below, views from the court on the appropriate use of SEF vary so greatly that the concept of variation might more descriptively be replaced by the concept of "randomness" were it not for the fact that there has been a consistent trend by the courts to accept SEF data as it is presented to them by institutions. In presenting an analysis and implications of these views from the court in relation to their validity, the detailed research literature on the validity of SEF will largely have to be bracketed. To do otherwise would take this article to far afield. Nevertheless, because the issue of validity is so central to this paper, an overview of the SEF validity literature is a necessary foundation for the following analyses.

There is a long and controversial research history on SEF, with most early reviews and extant opinion---though certainly not all---suggesting their general validity, with validity referring to the accuracy of SEF measuring teaching effectiveness. More recently, however, sophisticated statistical reviews of this past literature strongly suggest that earlier reviews of SEF literature were not rigorously analyzed and controlled methodologically, thus casting serious doubt on their validity. As Barnett (1996), Greenwald (1997), Greenwald and Gillmore (1996) demonstrate, past reviews have tended to not be sophisticated critiques. Positions, suggesting cautious support for validity of SEF while at the same time expressing concerns about the adequacy of their support, include, Abrami, Dickens, Perry, & Leventhal (1980). Reviews and empirical critiques that are critical of the validity of SEF include, Chacko (1983), Dowell and Neal (1982), Powell (1977), Snyder and Clair (1976), Vasta and Sarmiento (1979), and Worthington and Wong (1979). Some of the past reviews that have categorized the significant research that have found SEF to be essentially a valid measure of quality of instruction are: Cashin (1995), Cohen (1981), Franklin & Theall (1990), Holmes (1972), Howard, Conway, and Maxwell (1985), Howard and Maxwell (1980, 1982), Marsh (1980, 1982, 1984), Marsh and Dunkin (1992), and McKeachie (1979).

Cahn (1987) suggests that student ratings do not measure the instructional effectiveness or the intellectual achievement of students. SEF measure student satisfaction, attitudes toward instructors course, student personality, and their psychosocial needs. Cahn further suggests, students know if instructors are likeable, not if they are knowledgeable; they know if lectures are enjoyable, not if they are reliable. In a meta-analysis Cohen (1983)---who basically accepts the validity of SEF---concludes from his study, "While the magnitude of the average rating/achievement correlation for the thirty-three multisection courses is not overwhelming [14.4% of shared variance between ratings and the criteria], the relationship is certainly stronger and more consistent than we were led to believe..." (p. 455). And Dowell & Neal (1982) conclude that

> "The research literature can be seen as yielding unimpressive estimates of the validity of student ratings. At their most valid, then, validity of SEF refers to only 14% of the total variance. The literature does not, therefore, support claims that the validity of student ratings is a consistent quantity across situations. Rather, the evidence suggests that the validity of student ratings is modest at best and quite variable...The variability in obtained validity coefficients even in studies with reasonable methodological requirement lead us to suspect that the validity of student ratings is influenced by situational factors to such an extent that a meaningful, generalizable estimate of their validity does not exist. In general . . .no meaningful estimate of the validity of student ratings can be provided with confidence that is generalizable enough to be useful..." (59-61).

For example, studies demonstrate the following confounding variables: (1) Age, (2) gender, (3) class size, (4) year of student, (5) level of student, (6) instructor style, (7) subject matter, (8) major or elective course, (9) student interest in subject matter, (10) instructor grading difficulty, (11) anonymous v.s signed ratings, (12) whether students are informed of their use, (13) instructor present v.s instructor absent while completing the evaluation (see for example, Divoky and Rothermel, 1988), (14) length of class period, and a host of other variables.

Finally, the philosopher of science, Michael Scriven who has conducted rigorous work on evaluation procedures, (1995, 1993, 1991, 1988), particularly on the justification of inferring from ratings to conclusions about the merit of teaching on the basis of statistical correlations between ratings and student learning gains. He suggests that such inferences are invalid, unless a number of stringent conditions are met on the design, administration, and use of such ratings. He further suggests of faculty evaluation in general that, "All are face-invalid and certainly provide a worse basis for adverse personnel action than the polygraph in criminal cases. Based on examination of some hundreds of forms that are or have been used for personnel decisions (as well as professional development), the previous considerations entail that not more than one or two could stand up in a serious hearing." Given this highly questionable state of affairs on the validity of SEF, the question is how do courts view validity in relations to its use for administrative purposes?

The Courts' Approach to the General Accuracy and Psychometric Validity of SEF [5]

An issue directly related to the reliance on SEF for administrative purposes is its validity. Presumably the more valid SEF data in a given case, the more justifiable is the reliance on it for administrative purposes. From the legal cases reviewed (in Haskell, 1997b), it is clear that the courts tend to accept SEF data as presented to them by institutions.

> *Summary:* With regard to requiring the general and statistical accuracy of SEF, legal reasoning and rulings can be summarized (see Haskell, 1997b) as ranging from: (1) accepting statistical analyses as a part of a plaintiff's effort to establish discriminatory treatment if it reaches proportions comparable to those in cases establishing a prima facie racial discrimination, (2) cautioning that statistics are not irrefutable, with their usefulness depending on surrounding facts and circumstances of a case, (3) maintaining that the court need not consider validity and is under no obligation to establish the accuracy of administrative interpretations of SEF, (4) that tenure criteria are not drawn with "mathematical nicety," (5) administrator's failure to perform statistical comparisons is not arbitrary and is reasonable, (6) especially if such is not required by a Faculty Association Contract, (7) nearly any use made of SEF, regardless of its validity, is acceptable if it followed the standard practice of the university, (8) that creativity, rapport with students and colleagues, teaching ability, and other qualities are intangibles which cannot be measured by objective standards.

Some courts (e.g., *Fields V. Clark University*, 1987) have noted even when SEF are not gathered and evaluated according to accepted standards of scientific polling procedures it is nevertheless acceptable if the process followed standard practice involved in other tenure decisions at the university (p.671).

While there does exist a "substantial evidence" standard which gauges whether an institution's decision-making body carefully considered the evidence and had a substantial body of evidence on which to base its decision, and an "arbitrary and capricious" standard which gauges whether a deciding body acted without reason or irrationally, (See Kaplin,1995, section 1.4.3.6. Standards of Judicial Review and Burdens of Proof 35), it appears these standards are frequently ignored in relation decisions based on SEF.

In general, the exception to the courts almost total disregard for the validity of SEF has been in cases involving EEOC issues. In such cases, the courts require precise accuracy. I will address this issue in more detail in the section of disparate treatment and impact below.

Historical Overview of the Courts' Approach to the Validity of Faculty Evaluation Data

As noted previously, unlike general performance evaluations of faculty, SEF does not have a categorical legal history. Since SEF is but a subset of faculty performance evaluation in general, it is appropriate to briefly review the history of this more general area. Given SEF as a subset of faculty performance in general, it is accordingly not surprising to see that the view from the courts on the validity of SEF parallels that of the courts view of faculty performance evaluation.

Historically, in terms of faculty evaluation instruments in general, (on both secondary and postsecondary levels) it is widely agreed by legal scholars (Baez, Benjamin, and Centra, 1995) that "Despite the subjectivity of measuring the quality of a faculty member's scholarship, service and teaching accomplishments, courts will rarely, if ever, question the appropriateness of an institution's criteria (or how they measure them) for granting reappointment, promotion, or tenure....they will rarely substitute their judgments for those of peer review committees....Although juries may have less deference" (p.139). It might also be added that courts will seldom question administrative judgements of evaluations. It seems that faculty who challenge institutional evaluation tools very rarely succeed. Although the legal "competent and substantial evidence" standard places a significant burden of proof on the educational organization, it has not generally required that faculty assessment instruments are professionally validated (Rebell, 1990; Kaplin and Lee, 1995). Such rulings do, however, appear to vary by state or federal jurisdiction.

Psychometric standards of validity, reliability, and specific evaluation techniques, are rarely incorporated in state laws, regulations, or common-law standards. Accordingly, cases that involve evaluation have tended to focus on adherence to specific procedural requirements as set forth in state law or on general common-law notions of fairness and due process, not on expert psychometric standards. Although state courts will require strict adherence to the procedural aspects of these requirements and will strike down an arbitrary failure to use any apparent evaluative criteria, the state courts tend not to probe the substance of evaluation criteria or methods (Rebell, 1990; Kaplin and Lee, 1995). As Copeland and Murry (1996) have put it, "the judiciary has generally behaved as though it believed that evaluations were made only after careful deliberation and with procedural due process protections. In short, the judiciary has tended to act as if colleges and universities could be trusted to act in good faith" (p.246).

Rebell (1990) outlines what he describes as a "striking example of the courts' traditional deferential attitude toward teacher evaluation" data (p.337). The decision of the United States Court of Appeals for the Eighth Circuit in Scheelhaase v. Woodbury Central Community School District (1973), involved the dismissal of a teacher whose contract had previously been renewed over a ten-year period. The reason for her termination was that she was incompetent as indicated by the low scores of her students on the Iowa Test of Basic Skills (ITBS) and the Iowa Test of Educational Development (ITED). Despite the a number of expert witnesses testifying that it was inappropriate to use such test scores as a basis for evaluating a teacher performance, the court dismissed Scheelhaase's claim. The claims were considered basically irrelevant by the court because "such matters as the competence of teachers and the standards of its measurement" are not matters of constitutional dimension.

This early case involving a public school teacher is significant both because (a) of the Court's apparent lack of concern with the serious psychometric issues raised by a reliance on student achievement scores as a sole stated basis for termination and (b) because of the Court's almost total reliance on a school administrator's psychometrically unsubstantiated, and quite possibly equally erroneous evaluation. One of the concurring Scheelhaase case judges bluntly stated:

> The Board was entitled to rely upon the recommendation of conclusions of its superintendent, not-withstanding the existence of strong opinions contrary to his regarding the use of the ITBS or ITED tests as a tool of Leacher evaluation...Thus, its decision, even though premised upon an apparently erroneous 'expert opinion 'cannot be faulted as arbitrary and capricious. The Board's mere mistake in judgment or in weighing the evidence does not demonstrate any violation of substantive due process. (Emphasis added).

Thus, even when states use student achievement scores as an index of faculty proficiency, [6] courts have had an "apparent lack of concern with serious psychometric issues raised by reliance on student achievement scores as a sole stated basis for termination," again, relying on administrator's unsubstantiated evaluations (Rebell, 1990). Thus, courts have historically adopted the position that they are not qualified to second guess peer-review committees, at least as long as committees do not act arbitrarily and instruments are consistently and fairly applied (Baez, Benjamin, and Centra, 1995; Kaplin and Lee, 1995; Rebell, 1990). Traditionally, notes Rebell, most other courts have tended to take a similar deferential stance in teacher evaluation cases.

There seems to be two exception to this. The first is in discrimination cases. In general, courts have tended to only require precise accuracy in cases where EEOC issues are involved (See below). The second, is in claims of unfair treatment because of exercise of First Amendment free speech rights, including union-organizing activities, or allegations of denial of Fourteenth Amendment rights to due process by tenured teachers or others with a reasonable expectation of continued employment will trigger federal court jurisdiction with greater scrutiny of data (Rebell, 1990).

Acceptance of Administrative Subjective and Untrained Evaluator Judgements Of SEF Data [7]

An issue directly related to both the reliance on and statistical accuracy of SEF are views of the court regarding accepting or not accepting subjective administrative judgements of faculty teaching effectiveness.

> *Summary:* With regard to accepting the subjective judgements of administrators evaluation of SEF, the legal reasoning and rulings can be summarized as ranging from: (1) accepting administrative subjective judgements if (2) they are deemed sincere (3) grounded on some evidentiary basis (4) if made on the "vigor and variety of student criticisms" (5) "not arbitrary or capricious and were exercised honestly upon due consideration," (6) based upon "much experience in reviewing student evaluations, (31) reasonably draw on that experience (7) and have ruled that Presidents are not bound by factual findings made by majority members of a faculty.

Not only have the courts not traditionally examined faculty evaluations rigorously, they have tended not to require that evaluators be trained in the use, analysis, and interpretation of evaluation instruments. In general, state courts reviewing teacher evaluation practices will not analyze directly the substantive criteria used to evaluate teachers, nor the or qualifications of the raters. (Rebell, 1990). There are exceptions, however.

Some states, like Florida and Pennsylvania now mandate such training. Florida specifically mandates school boards to provide training programs to "ensure that all individuals with evaluation responsibilities understand the proper use of the assessment criteria and procedures" (Fla. Educ. Code, /sec 231.29(2). In Pennsylvania (Rebell, 1990), employees must be evaluated "by an approved rating system which shall give due consideration to personality, preparation, technique and pupil reaction in accordance with standards and regulations for such scoring as determined by rating cards to be prepared by the Department of Public Education...." (p.345-6).

SEF as Social Judgement and Diagnosis

Given the courts assumptions regarding validity and the untrained judgement of those making decisions based on SEF, a part of influencing the courts is demonstrating relevant research. In the research on social judgement and clinical diagnosis, it is clear that the manner in which nearly all SEF data are analyzed is but a subset of the social judgement and clinical diagnosis literature, involving the same logical and cognitive bias and distortions that result in the pervasive inaccuracy of social judgement in general and clinical diagnosis in specific. The findings of the judgement research literature applies to students making such judgements in evaluating faculty and to those interpreting the results; they are in fact making diagnoses.

Psychological research has recognized the severe cognitive problems and limitations of "intuitive," and "experience-informed" everyday judgements for over thirty years, (Dawes, Faust, and Meehl, 1989; Faust, Guilmette, Hart, Arkes, Fishburne and Davey, 1988; Garb, H. N. 1989; Hayes, 1991; Larkin, McDermott, Simon, and Simon, 1980; Rabinowitz, 1993) yet the mistakes continue in everyday practice situations. Interpretation of SEF are no different. As two authors who consider SEF literature valid (Franklin & Theall, 1990)---point out:

> Even given the inherently less than perfect nature of ratings data and the analytical inclinations of academics, the problem of unskilled users, making decisions based on invalid interpretations of ambiguous or frankly bad data, deserves attention. According to Thompson (1988, p. 217) "Bayes Theorem shows that anything close to an accurate interpretation of the results of imperfect predictors is very elusive at the intuitive level. Indeed, empirical studies have shown that persons unfamiliar with conditional probability are quite poor at doing so (that is, interpreting ratings results) unless the situation is quite simple." It seems likely that the combination of less than perfect data with less than perfect users could quickly yield completely unacceptable practices, unless safeguards were in place to insure that users knew how to recognize problems of validity and reliability, understood the inherent limitations of rating data and knew valid procedures for using ratings data in the contexts of summative and formative evaluation. (79-80).

The authors conclude by noting, "It is hard to ignore the mounting anecdotal evidence of abuse. Our findings, and the evidence that ratings use is on the increase, taken together, suggest that ratings malpractice, causing harm to individual careers and undermining institutional goals, deserves our attention." (p.79-80). Recognizing such problems is not methodological nit-picking; they are pragmatic, paradigmatic, and scientifically fundamental.

Variables Affecting Validity Not Taken Into Account When Assessing SEF [8]

In conducting any research, it is a given there are a host of variables that affect outcomes. Put in experimental terms, there are a host of independent variables that affect the dependent variable (here teaching effectiveness). The question is, how have courts addressed this crucial issue that impacts so centrally on validity of SEF data?

Instructional Variables

Legal cases concerned with the validity of SEF occasionally note various instructional factors that were not controlled in the faculty evaluation process.

> *Summary:* The variables noted in the legal cases reviewed include, (55) not controlling for class size, i.e., those obtained in small seminars from those obtained in large lecture classes, (56) those obtained from tenured faculty from those obtained from non tenured junior faculty, (57) not performing appropriate comparisons of SEF with other faculty, (58) noting SEF in all courses, not just to problem courses, (59) not mistaking student 'response' figures for actual student enrolment figures when using them to determine student attraction to a course, (60) using all courses taught, (61) taking into consideration faculty teaching a wide range of courses, versus those with lighter teaching loads, (62) number of new courses taught in a year,.(63) whether graduate courses were taught at the same time as teaching undergraduate courses, (64) selectively mentioning only negative student comments, or (65) overly weighting negative comments, and (66) different procedures for gathering student opinion.

Courts sometimes weigh these variables heavily, in most cases, however, the courts either ignore them or do not weigh them very heavily in the total context of a particular case. [9]

Student Biases Variables [10]

A significant issue is how courts view student biases in assessing the reliability and validity of SEF.

> *Summary:* Student bias variables include reactions to (48) academically

> demanding faculty, that (49) thus thwart student expectations, (50) difficult examinations (51) tough grading policy, (52) heavy workload in a course. (53) While most courts ignore these student biases in SEF, (54) occasionally a court will recognize that difficult courses have to be given to the students and that such material is difficult for even the best teacher to get the material across.

In general, however, it is overwhelmingly clear that courts seldom take these variable into account, despite the fact that such reactions often function as generalized affective overlays on SEF (see below).

Popularity Variables and Effectiveness [11]

A related student variable issue is the extent to which SEF measures popularity, not teaching effectiveness. Accordingly, it is instructive to see how courts view this issue.

> *Summary:* Court rulings range from saying that (9) in cases of exceptional research faculty that popularity should not play a role in termination due to teaching, to (10) in normal cases that a measure of popularity is related to teaching effectiveness.

While not noted frequently, popularity appears to be generally assumed to be involved in teaching effectiveness. But again, the courts are mixed on this issue as well. In terms of the research literature there is little to no support for popularity being a measure of teaching effectiveness in higher education. [12]

The Courts' Reliance on Both Quantitative Data and Qualitative Comments in SEF [13]

Reliance on SEF v. Peer Evaluation

Is it considered acceptable, for example, to rely heavily or even solely on SEF, or must they be used in conjunction with other evaluative methods?

> *Summary:* From the cases analyzed, it can be seen that court rulings range from saying that (1) relying primarily or solely on student evaluations is acceptable, to (2) placing little exclusive reliance on SEF, (3) in rare cases SEF can not be permitted to stand in the way of promoting or retaining professors who are excellent in non teaching areas, (4) tenure decisions can not be based solely on SEF by students who have not been made aware of the ramifications of their evaluations, (5) anonymous documents or those "based on hearsay" should not be included in a faculty member's file, (6) students should be made aware of the purpose and ramifications of their evaluations of faculty, (7) anonymous student evaluations should not be used, (8) peer evaluations must also be a part of evaluating teaching.

Again, courts range widely on the exclusiveness or non exclusiveness of

SEF, even though books on how to conduct faculty evaluation (by authors who basically accept the validity of SEF, e.g., Seldin, 1984; Theall, and Franklin, 1990) for some time now have consistently emphasized that SEF should not be used as the only and/or primary method for assessing teaching effectiveness.

Numerical Ranking of Faculty [14]

An important issue is how the courts view the relative weighting of SEF in administrative decisions of teaching competence. It seems to be common practice to ordinally rank and compare faculty to each other according to average SEF numerical scores.

> *Summary:* From the cases reviewed, numerical scores from SEF often result in faculty (22) being compared relative to other faculty, (23) being ranked relative other faculty, (24) with distinctions often being made on the basis of tenths of a decimal, (25) with most courts accepting these fine decimal distinctions.

Despite the above overview of the research on the highly questionable validity of SEF, institution administrators and the courts continue to make and accept fine numerical distinctions in faculty scores from student evaluation questionnaires to ordinally rank faculty. Even given that SEF is valid to a level accounting for 14% of the variance, it is not psychometrically appropriate to accept such ordinal rankings.

It should be noted that SEF rate the majority of faculty as above average---whatever this means.

Ordinal scales do not tell us if a faculty half way down the scale is only half as good as the top ranked member. Thus without a criterion referenced standard, we have no way of knowing if everyone on the scale is an effective teacher, or conversely an ineffective teacher. Moreover, should all faculty who fall below the statistical "average" be eliminated? And if so, using the same logic, should we rank order and thereby eliminate all Olympic team members who fall below the team average? If the answer is 'yes,' then (a) we eliminate highly functioning athletes, and (b) it leads to an infinite regress where we end with only one or two on any given team. Currently, we have no idea if "statistical average" means good, bad, or indifferent teaching in terms of instructional effectiveness.

Use of Qualitative Written Student Comments [15]

Over and above quantitative data, the use of written comments, often *single instances*, by students on their SEF forms seems wide spread by both educational administrators, faculty evaluation committees, and the courts.

> *Summary:* For the use of student comments, court views ranges from (33) placing importance on a single comment (34) to several comments as

> significant information, (35) maintaining that statistical analyses of SEF need to be bolstered by individual comments, (36) maintaining that while some very negative---e.g., racist, sexist---comments may be found, the court may find that they do not render SEF unreliable, (18) that such instances or "*impressions*" may be validated after the fact, (37) negative comments often seem to outweigh positive ones, and (38) may often outweigh numerical data to the contrary, (39) negative comments need not be verified before acting on them, to (40) that negative comments can not be used to undermine otherwise generally favorable comments received in an annual performance review.

Clearly the views from the court suggest the legitimacy of not only using what is in fact anecdotal data, but often to raise it above more systematic (averaged) data.

Mixed Student Comments [16]

Just as quantitative SEF data may be bimodal, so too written student comments may also be bimodal or mixed. How do courts (indeed, educational administrators, and faculty evaluation committees) view and pronounce on such data?

> *Summary:* With regard to non numerically assessed written student comments, they are often qualitatively characterized as (41) a few were ambivalent, (42) a considerable number, (43) of mixed result, and selectively recognized: (44) it would only be fair to add that there were a number of comments in favor, (45) there were also some negative comments, (45) sometimes placing the greater weight on past evaluations of teaching over current comments, (47) sometimes placing greater weight on current comments over past positive evaluation of teaching.

Again, with regard to single and mixed comments on SEF, the courts (administration, and faculty evaluation committees, See Appendix) tend to weigh them far above their non representative and anecdotal-data value.

It seems to be generally assumed by most faculty and administrators that SEF are used by virtually all schools in the U.S. It is further assumed by many that SEF is necessary for both faculty evaluation of teaching effectiveness and thus for quality control of student learning. While its use is clearly wide spread (see Seldin, 1984; Crumbley, and Fliedner, 1995) in the U.S., and is increasing in Europe (Husbands, and Fosh, 1993), what is not generally recognized is that there are schools that preclude its use in salary, promotion and tenure decision either totally, or in part, by precluding the use of qualitative students comments.[17]

Transcendent Value of a Professor Over Teaching Quality [18]

Despite the importance placed on teaching, there is precedent for both school policy and the courts---under certain conditions---to ignore poor teaching as indicted by SEF.

> *Summary:* (11) The courts and educational administrations can not allow low SEF to stand in the way of promoting or retaining professors who may be

> world renowned scientists, (12) deemed nationally or internationally exceptional as a researcher, courts may nevertheless disregard SEF, (13) at least in these two cases the courts did not find the faculty exceptional. It would be interesting to see if what the court seems to accept in principle exists in fact.

The above collective categories abstracted from court cases are illustrated by a denial of tenure case described in the Appendix below, by a (non litigated) case that contains an interesting difference from most of the cases reviewed here.

Procedural, Burden of Proof, and Policy-Decision Criteria in Assessment of SEF

Other overlooked issues involving SEF and its validity are the problems of (a) content versus process, i.e., whether the assessment of SEF data constitutes a process or procedural issue or (b) is simply a content issue.

Validity Assessment of SEF as Procedural or Process Issue

An exemplar of the content and the procedural/process distinction is often exhibited between trial and appellate courts. The latter often only judge if correct procedural/due process was followed by a lower court. The content v procedural/due process distinction is typically used by college campus grievance committees. When a tenure committee, for example, renders an unacceptable decision, a faculty member may challenge the decision. A grievance or appeal committee then may review the decision only in terms of if the correct process or procedures by which the decision was made was followed. The point here is that many such appeals committees do not define looking at the procedures by which SEF data were gathered and analyzed by a tenure committee or administrative evaluator as procedure/due process (e.g., whether the tenure committee just 'eye balled' the data and student comments, whether they compared the data to other similar faculty SEF, etc.), but as content and therefore not within its purview. Grievance committees often therefore will not review the substantive content of SEF data on grounds that it is not a procedural or process issue.

In general, given the courts tendency to accept the validity of SEF data, at least by default, how SEF data are assessed and used is often considered to not be a process/procedural issue. At least one court has, however, considered how SEF data is assessed and used as procedural. This is evidenced in *Christopher Turner v. The President of the University of British Columbia* (1993), where it was stated that

> the Dean said, "there were few students in undergraduate literature courses since 1986/7---(3,8, and 6 respectively," thus mistaking student 'response' figures for actual student enrolment. The Board concluded that (5) "This misunderstanding is in our opinion sufficient in itself for a reconsideration, since teaching was the focus..." (p.3), and (7) "we think that the comments and emphasis on the size of Dr. Turner's classes as evidence of poor teaching

> are open to objection and *constitute errors of procedure and/or evidence*" (p.6). [italics added]

As noted above, however, it appears that most courts, and indeed, perhaps most faculty grievance committees (See Appendix below) have not considered how SEF data is analyzed as a procedure/due process issue. The issue of the validity of SEF, then, would appear to have legal "due process" implications.

Decision Criteria and the Scientific Precautionary Principle

Since SEF has haphazardly evolved along with a general acceptance of its validity as an appropriate measure of faculty teaching effectiveness, the burden of proof somehow has been placed on faculty-as-challengers of such data to scientifically prove that SEF data is *not* valid--- a strange state of affairs, at least in science. And the standard of proof required has been typically high. In effect, faculty are guilty until proven innocent. So the process that exists is:

1. Either (a) a legal abdication of the assessment of SEF by the court, relying on the good faith evaluation of SEF data by the institution, or (b) the court simply assuming its validly.
2. Placing the burden of proof on faculty who challenge the data of demonstrating with scientific levels of certainty (statistical significance or confidence level) that the data is not valid.

Given---at the very least---the controversial assessment of the validity level of SEF in measuring teaching effectiveness, in terms of decisions and policy perhaps we should err on the side of caution in applying such data for administrative purposes. In the field of environmental science, Lemons (1996) and Lemons, Shrader-Frechette and Cranor (in press) have suggested a Precautionary Principle when making policy decisions. In essence, this principle says that when making policy decisions about environmental harm, given (a) a certain level of possible harm, (b) the complexity/uncertainty of data, and (c) the high level of proof (typically a 95 per cent confidence level) required for a scientific finding to be accepted by scientists, setting policy should not be based on this level of scientific proof. The reason is this: To wait for such a confidence level may be too risky given the level of harm that may be indicated (by the existence of data with a lesser confidence level suggests). In short, using scientific criteria that have been adopted for doing science may often not be appropriate criteria for making policy decisions.

The reasoning surrounding the Precautionary Principle is too complex to fully delineate here. The reader is referred to the citations. In the meantime consider the following analogy that in broad outline exemplifies the spirit of the Precautionary Principle: A dangerous tiger has escaped from a local zoo a few miles from your house. In the back of your house is a wooded area. Your child wants to go out and play in the woods. No one has actually seen the tiger in the woods or anywhere else around the neighborhood. In other words, there is no scientific level of evidence that the tiger is anywhere around, or that your child would be in immediate danger by playing in the woods. Do you let your child out to play in the woods?

In most areas of science, the rule is to avoid type-I error---asserting there is an effect when there is none, and therefore place the burden on those who postulate an effect rather than on those who postulate no effect---and not so much concerned with avoiding type-II error---asserting no effect when there is one. In adopting SEF data as indicating teaching effectiveness administrators, faculty evaluation committees and the court have engaged in type-I error---given both the level and burden of proof.

Now there are two implications for the Precautionary Principle as applied to SEF in relation to faculty and instructional quality. First, given (a) the haphazard way SEF have been introduced and accepted by the courts (b) the level of possible harm of accepting SEF for administrative purposes of salary, promotion, denial of tenure or non reappointment, to that faculty and more importantly (c) the effects of SEF used for such purposes has on the quality standards of higher education (see Haskell, 1997a) should such a burden of proof be demanded by the court of faculty challenging SEF data? Certainly, as shown below in disparate treatment and disparate impact cases, a kind of Precautionary Principle is already in effect. Second, given the at least clearly conflicting evidence of whether SEF demonstrates teaching effectiveness of a faculty, should not administrators and faculty evaluation committees apply, for the same reasons, a similar Precautionary Principle stance?

The Court's Approach to Validity of SEF in Relation to the Principles of Disparate Treatment and Disparate Impact [19]

Given the above findings on how the courts have tended to treat SEF validity issues, I would now like to further look at the implications. Federal courts---and to a lesser degree state courts---have adopted a more stringent approach to testing teacher evaluation cases, at least regarding primary and secondary teachers. According to Rebell, the four main reasons for this change are (1) the wider use of more stringent evaluation techniques by institutions, which largely stem from legislative reform initiatives that have led to an increased number of denials of teacher certification and terminations, (2) a disproportionate number of these certifications and termination involve members of minority groups, (3) legal developments have broadened the jurisdiction of the federal courts to consider issues of social reform, and (4) judges' own increased experience in assessing psychometric techniques in employment discrimination cases. It is perhaps 2 and 3, however, that have had the most impact on the courts (Kaplin and Lee, 1995; Rebell, 1990).

Educational reform issues from desegregation, special education, and other school-based litigations, has made the courts more experienced and more inclined to scrutinize educational testing requirements. As the consequence of federal Equal Employment Opportunity Commission (EEOC) criteria, in today's civil rights climate, courts are more likely to scrutinize the validity of the faculty evaluation instrument, especially in terms of racial, gender, and age discrimination.

Disparate Impact

In regard to teacher evaluation in general in cases involving claims of discrimination under the equal protection clause of the Fourteenth Amendment, or under the anti discrimination statutes enacted to protect members of racial and ethnic minorities, women, handicap conditions, age, and other protected groups scrutiny of the case tends to be more probing and stringent. Such cases are of two basic kinds: (1) those involving discriminatory intent, called disparate *treatment* claims, and (2) those involving no intent, called disparate *impact* claims (see Kaplin and Lee, 1995, section 3.3.2.1.).

Disparate impact claims in personnel evaluation is the use of assessment procedures that are facially (on their surface, or methodologically) neutral in their treatment of different groups, but which produce evaluation outcomes that inadvertently fall more harshly on one group than on another. Thus, proof of a discriminatory motive is not necessary to establish a disparate impact claim. To establish a *prima facie* case of such adverse impact, a minority need only show a causal connection between the facially neutral employment practice and the disproportionate negative or adverse effects on him or herself as a member of a protected group. For example, a university tenure process may be found to discriminate against females because the evaluation process or evaluation criteria favors male faculty more than female. In such cases, rigorous statistical analysis is typically used to establish disparate impact.

Discriminatory treatment and disparate impact claims has made courts more inclined to specifically analyze educational testing instruments for validity, and this increased involvement by the courts is predicted to increase. As of 1990, 41 states have mandated some form of standardized testing requirements as part of their teacher certification process Because many of these exams are claimed to have a disproportionate negative impact on minority candidates, competency tests have triggered a number of large scale federal class suits. Again judge Rebell (1990) notes,

> In June 1988, the United States Supreme Court issued a ruling which is likely to accelerate the trend toward increased judicial involvement in teacher evaluation matters. That case, *Watson v. Fort Worth Bank and Trust* (1988), extended to judgmental employment practices the Court's 1971 holding in Criggs v. Duke Power Company (1971) that standardized employment tests having a **disparate impact** on minorities must be shown to be job-related. Although the Court's ruling in Watson was unanimous, there was substantial disagreement among the Justices as to how closely courts should scrutinize particular practices and **validation techniques**. Whatever the precise standard of review ultimately implemented, there is little doubt that the federal courts will be more likely to scrutinize nonobjective evaluation procedures as a result of Watson (p.339).

Thus any instrument or evaluation criteria that *in effect* places an unfair burden on those being evaluated has been judged to exhibit what is legally termed *disparate impact*.

The present point is that while the courts have not, and continue to not rigorously scrutinize SEF, they have for sometime now applied fairly rigorous standards to evaluations both in the workplace and in academia to cases involving discrimination of protected groups, whether the discrimination is purposeful, or by disparate impact. Not every indication of racism, however, may be considered by a court to be proof of discrimination.

For example, In *Yu Chuen Wei v. Vermont State Colleges Faculty Federation* (1995), the Labor Relations Board said, "with respect to comments that while some students had written that she was a "slant eyed bitch," and that she should "go back to China....We also are not persuaded that the racism evident in the student evaluations of Grievant made student evaluation results unreliable. The percentage of evaluations in which racism by students was evident was approximately one percent of the total evaluations" (p.306). Assuming some level of covert racism, how does one disentangle the generalized affective racist and sexist overlay of students evaluation on a total questionnaire?[20]

The Disparate Treatment and Impact Principles Generalized

In the evolution of any legal policy or principle its extension often occurs by generalization or analogical transfer, extending a principle thought to apply to only one area to other areas (See for example, Anderson and Schadewald, 1991; Golding, 1984; Levi, 1949; Marchant, Robinson, Sunstein, 1993). Currently both corporate and academic cases of straight forward discrimination and the more inadvertent discrimination cases based on disparate impact often trigger the courts to rigorously scrutinize the methodology and statistical data of such evaluations not typically accorded to non discrimination cases. Presumably, in discrimination cases the court's interest is in establishing validity and using rigorous statistical methods in ascertaining the "truth." If this is the case, then by clear logical implication and inference---as we have seen---in generic cases of evaluation the court could be said not to be in the truth business. As documented above, in non discriminatory cases, courts have assumed the "truth" lay in the appropriateness of an institution's criteria and rarely would substitute their judgments for those of peer review committees, adopting the position that they are not qualified to second guess peer-review committees, at least as long as committees do not act arbitrarily and instruments are consistently and fairly applied. The burden of proof is on the faculty challenging an institutional decision. Some courts have only been concerned with consistency and fairness of application, even if the methods of evaluation are clearly defective.[21] Generally, however, the courts have acted as though they believed that institutional evaluations were made only after careful deliberation and with procedural due process protections.

The question is: why not make the same assumptions regarding discrimination? The answer is that, understandably, the courts have accepted that there has existed a widespread conscious and non conscious ethnic, gender, age , religious belief, sexual orientation, and handicap bias in society, such that they can not simply rely on the "truth" or good faith behavior of an institution or its data. Given this, the argument is made that herein lies the distinction, and reason, for treating non discriminatory cases differently from cases where either discrimination has been charged (treatment claim) or where discrimination has be inadvertent (disparate impact claim).

Thus the courts have tended to accept the judgement and "good faith" motivations of organizations. Unlike in the past, however, just as the data are in regarding discrimination of protected groups in academia, so too the data are now sufficiently in to cast serious doubt on the courts assumption of "truth" residing in corporate and academic data on discrimination, so too is it in on (a) the questionable validity of SEF, (b) the internal politics of administration and faculty relations which can revolve around student retention and unpopular ideas, and (c) the economic pressures on institution to not tenure faculty and to sometimes terminate tenured faculty, all of which can have serious contaminating consequences institutional decisions.

The importance of this is that while courts have scrutinized SEF for evidence when civil rights discrimination has been questioned or suspected, they have not applied the same rigor to the validity of evaluation instruments or have held as suspect other institutional biasing variables. The courts continue to assume a kind of pre 1960s academic Camelot. If such a round table of academic knights ever did historically exist or was merely mythical, it certainly now exist only in myth.

One compendium of legal findings in higher education specifically notes SEF and recognizes the accepted application of the principles of disparate treatment and impact along racial and gender lines in SEF. It should noted that the disparate treatment and impact issues when applied to SEF, is of course, no different than any other disparate impact case, except that the student evaluation data in such cases will be scrutinized by the courts. The authors (Baez and Centra, 1995) suggests that the SEF research in the area of race and gender discrimination, has been inconsistent, and suggest that while deserving of more attention, the inconsistency of the research makes it unlikely that the courts will sustain such a claim. Some courts, however, have found in favor of faculty in such cases. For example,

> In *Cynthia J. Fisher v. Vassar College* (1995), after a bench trial, the district court found that, in denying Fisher tenure, Vassar had discriminated against her by reason of (a) her sex in violation of Title VII of the Civil Rights Act of 1964, (b) her age in violation of the Age Discrimination in Employment Act. The court found that the termination of Fisher's employment resulted not from any inadequacy of her performance, qualifications, or service, but rather from pretextual and bad faith evaluation of her qualifications. Scrutinizing Vassar's report on Fishers teaching ability which included reviews of her student evaluations that were said to reflect "consistent problems with clarity and her ability to illuminate difficult material" but which were otherwise generally positive. The district court found that the Vassar's biology department had distorted her teaching recommendations by "selectively exclud[ing] favorable ratings," by selectively "focus[ing] on the two courses in which she had difficulties" and by "applying different standards to her than were applied to other tenure candidates" (*Id.* at 1209). The court further observed that "the males tenured while Dr. Fisher was on the faculty were praised for their fine teaching while Dr. Fisher was criticized, although the facts on which the Committee's determinations were based (student

> evaluations, Biology Majors Reports and [Student Advisory Committee] reports) revealed that Dr. Fisher's evaluations were superior to theirs" (*Id.* at 1211). The court noted that statistical analysis may be a part of a plaintiff's effort to establish discrimination under a theory of disparate treatment.

The point here is that if this had not been a disparate treatment discrimination case the biases and distortions of data about her teaching student evaluations would likely have gone un examined. [22]

It would seem, then, that this discrepancy in the discrimination-based search for "truth" should be used as---and provide justification for---a kind of generalized disparate impact principle to legally invoke or generalize a fairness principle that the same rigor be applied to non civil rights cases such as SEF. As Kaplin points out, however, current law generally prohibits courts from such generalization. [23] So the issue of change apparently becomes not so much one for the court as it is a policy issue for both higher educational administration to use as a guideline and for legislatures to legislate change.[24]

Age discrimination in SEF is another possible bridge in this potential extension of disparate impact. The Age Discrimination in Employment Act of 1967 (ADEA) requires employers to evaluate persons on their qualifications or ability to competently perform their job, and not on the basis of age. Like any other employer, colleges and universities are likewise prohibited from considering a faculty member's age in making decisions about employment, salary increases, promotion, tenure, and retention. Yet, there is evidence that SEF do discriminate on the basis of age, with older faculty receiving lower student ratings (Feldman, 1983). There are a host of other variables like class size, or teaching a courses within a student's major as opposed to elective course, or teaching freshman v.s upper level students, that also make a kind of default "disparate impact" if such variables are not controlled in the analysis of SEF data.

What is being suggested here is that in the interest of justice, equity, truth, and in "fact finding," the courts and institutions should scrutinize all SEF data as rigorously as they do disparate treatment and disparate impact cases. Currently data and conclusions from SEF are seldom scrutinized (as indeed are other issues in the denial of tenure or promotion not equally scrutinized) as they are in discrimination and disparate impact cases. Justice, however, is not only blind to ethnic, gender, age, sexual orientation, religious belief, and handicap status, it is blind to institutional economic pressures and other biasing variables within academic institutions. Thus, biases and distortions of the SEF data are not revealed in non discrimination cases as they are in disparate treatment and disparate impact cases. As a consequence, in terms of revealing unfair attributions based on SEF data, those covered under EEOC guidelines have a "truth finding" advantage over those who are not covered.

Beyond Statistical Significance of SEF Research

Having reviewed SEF cases and examined the significance of validity, I would now like t to turn the issue of validity on its head. Underlying statistical research on SEF that attempts to establish its validity is a complex of contextual variables and assumptions seldom addressed.[25] In this section, I will address some of these contextual variables and assumptions that I suggest cut through and render the best of statistical research on SEF showing teaching effectiveness nearly irrelevant. Understanding is not acquired by statistical significance alone. Certainly showing statistical validity of SEF is a necessary condition, but it is not a sufficient condition for understanding their meaning and for its use in administrative decisions. It is an understanding of these contexts and assumptions that underlie statistical validity research on SEF that educational policy-makers and the courts need to think long and hard about accepting SEF for assessing instructional competence and using it for promotion, tenure, and reappointment decisions.[26]

Assumption # 1: Statistical Significance of Indicators of Teaching Effectiveness

An assumption underlying statistical analyses of SEF is that we know what the indicators of effective teaching are. To my knowledge, the research does not support this assumption. What makes us so sure that many of the questions we ask on SEF questionnaires are all that related to effective student learning. Consider, for example, the typical question "Was your instructor organized?" This question in turn entails a myriad of assumptions and conditions about effective teaching. Would Socrates, for example, be *perceived* as organized by most students---being peripatetic and just asking a lot of questions? And what makes us think-----at least for some students and some kinds of subject matter---that just going into class, being Socratic, asking a lot of provocative questions, and confronting students by challenging their belief systems may not be the most effective instructional and learning method in the long run to get students engaged and to think critically?[27] What evidence is there that either being perceived as organized or actually being organized is a necessary condition for effective instruction? I know of no rigorous supporting evidence. Indeed, many of my friends in the humanities, much to the dismay of my behaviorist and cognitive colleagues---and sometimes myself---would suggest that systematic and sequentially structured teaching methods are simply structural analogues of our technological society (see, for example the classic by Jacques Ellul (1964).

Consider, too, a question that, while it is not directly asked on SEF questionnaires is implied in other questions in various forms, inquiring "Does your instructor mainly lecture?" Though there is precious little rigorous evidence showing that lecturing is inherently an ineffective teaching method, it is clearly *persona non grata* among many educational theorist. Lecturing is "out" while collaborative learning is "in"---but apparently not so considered by many faculty (for both valid and invalid reasons).

While I happen to agree that being organized is generally good, and that collaborative learning is perhaps good for certain student populations, subject matters and desired outcomes, the question is: are they appropriate indicators of effective teaching applied to individual faculty as claimed? The answer to this question is they are not appropriate indicators of effective teaching applied to individual faculty---and this applies even if the statistical research strongly supported the claim. This is an important point that, as I recall, is addressed in the faculty evaluation literature only by Scriven (1988). I will quote Scriven at some length.

Scriven observes that in the attempt to render teacher evaluation more scientific the field rushed into focusing on research-based indicators, teaching indicators which sound research supposedly demonstrated are positively correlated with successful student learning. These indicators or

> Popular envies are structured presentations, active involvement, emphasis on positive reinforcement, high eye contact, high frequency of question asking, provision of learning objectives, frequent feedback, use of multi-media (p.4)....the provision of a brief outline of topics to be covered in a day's lesson can be justified on administrative grounds, since substitute teachers must get some guidance; but the requirement that anything like that be provided to students, for pedagogical reasons--a claim often said to be supported by research---cannot be justified. The use of instructional objectives or any other kind of advance organizer is simply a characteristic of ones style of teaching, not a duty of the teacher. Nor can such an outline be required as evidence of preparation (arguably a duty), since a teacher using a textbook--or for that matter, memory--may do as well or better than one with lengthy lesson plans listing activities and testing procedures (p.7-8).

The presence or absence of these factors, says Scriven, defines a style of teaching. He maintains that any reference to a 'teaching style' in teacher evaluation is not valid, regardless of whether there exists a research basis for thinking the style is correlated with teaching effectiveness.[28] He goes on to explain:

> A major source of confusion in discussing the use of indicators is that the research is often presented as showing that 'the best way to teach' is by using high eye contact (or whatever), whereas all it really shows is that there a slight tendency for better teachers to exhibit this characteristic, for reasons which might include the fact that they were taught to use it, although in fact it's not a help at all. The reader is seduced by the relative plausibility of the style recommendations, whereas you'd never buy the idea of using eye color or skin color. But plausibility isn't necessity, and absent necessity, you're just a stylist Our kids don't need stylists, they need good teachers; and if you can't distinguish the two, you're in the wrong business (p.7).[29]

Scriven is not denying the validity of statistical inference. Useful information is contained in a statistical correlation, and there are circumstances in which that information can be put to good use. It can even be put to good use in making decisions about people---but only when no better data is available because of limitations on time or resources. [30] Scriven maintains that such teaching effectiveness indicators are invalid

> for essentially the same kind of reason that the evaluation of personnel by the color of their skin or their church affiliation is necessarily invalid. While it is true that much racial prejudice, sexism, etc., is based on false beliefs about the groups discriminated against, the essential flaw in it goes deeper than that. The essential flaw is that even if women in general are less strong than men, you shouldn't. use gender to discriminate against a particular candidate for a position as a luggage-handler, but only a job-related strength test or series of observations in a trial period on the job. And this is nor just for ethical/legal reasons, but also for scientific reasons and reasons of efficiency (p.4)....Which means you can't discriminate against a teacher on the grounds

> that s/he exhibits some approach to teaching that research has shown is less likely to be successful. Whites are statistically less likely to be good basketball players than blacks, but you can't kick the whites off the squad the day you discover that the statistics are worse than you thought. nor would you be any good as a coach is you used skin color as a criterion for selection. You have to look at the individual's success, not at the success of groups to which the individual belongs (p.4).

Finally, Scriven suggests a reason for the almost total disregard of the validity of SEF by the courts documented in this paper (and my previous paper, Haskell, 1977b). He understands the implications for courts recognizing the fallacy of such indicators: He says, The current fallacy of using such statistical-indicators are,"as certain to crash in the courts---eventually---as the most blatantly racist hiring practices. We may have only a short breathing space before the courts and defense attorneys begin to see the underlying similarity of these two approaches....The consequences for states and districts will be chaotic; old decisions may be reversed on appeal, huge damages may be awarded, those hearings will clog the system, and there will be no legitimate process to take the place of the illicit one (it is because of this potentiality for disaster that we are giving a longer than-usual treatment of the issue here) (p.5).

Assumption # 2: Statistical Significance of SEF of Teaching Effectiveness Measures Appropriate Learning

An assumption that is virtually unnoted in the literature is that given SEF is eventually found to measure teaching effectiveness---and this "given" is only for the sake of the current argument---it is assumed that what is thereby being measured is appropriate learning. This assumption is arguably incorrect for at least two reasons. I say it is arguably incorrect, as whether the assumption is correct or not depends on other differing assumptions about higher education.

First, let us not fool ourselves into thinking that we know what effective teaching is for all populations of students and subject matters. There is no shortage of possible indicators of effective instruction and learning, but most are not articulated within an adequate theory of effective instruction or learning. At the very least, "effective" is relative to a given student population. And when referring to teaching effectiveness are we referring to measuring short term or long term learning?[31]

In addition, as Abrami (1989) and others (see Cohen, 1983) have suggested, most studies on the relationships between student ratings and instructor-generated student learning have been done with learning outcomes collected largely from freshman classes, and---more importantly---learning at the lowest level of Bloom's taxonomy. Similarly, the literature on transfer of learning shows that when student transfer of learning is found, it reflects the lowest level of concrete transfer. So even if we are effective in achieving this level of effectiveness, what have we achieved? This brings me my main point.

I suggest that teaching effectiveness and appropriate learning in higher education are two different logical and empirical entities. I shall now address these two differing assumptions together. If the data showing

(a) student level of unpreparedness, (b) student ability level as measured by most national tests, (c) unrealistic student expectations about learning, (d) grading, (e) feeling of entitlement, (f) motivation level, (g) good faith motivation for evaluating faculty, (h) maturity level, and (i) hours spent studying have been either in decline for years, or have become increasingly inappropriate is accepted, then effectiveness in teaching most of these students does not necessarily---and most likely does not---mean appropriate learning. For purposes of clarity (and at some risk of seeming not only insensitive, but as a right wing radical, which I assure the reader I am not), let me demonstrate why teaching effectiveness is separate from appropriate learning by using what may be considered an extreme scenario as an example: Suppose that the American Disabilities Act as applied to higher education is amended to include having to admit the mentally retarded, thus requiring making whatever instructional adjustments need to accommodate their disability.

Now assume that such adjustments are made, e.g., speaking slower, simplifying and otherwise decreasing the amount of content to be mastered, along with the depth of understanding and critical thinking. In addition, assume that if such adjustments and other classroom behaviors that were once apporpriate for a previous level of student are not accommodated and that this is reflected in low SEF score. Now assume that because of pressures such adjustments have been made and that SEF findings for those teaching the disabled students unequivocally shows teaching-effectiveness. The question then becomes: is this appropriate learning for a higher education course? [32]

Most will likely respond to this question with a resounding "no." Some, on the first assumption noted above may say "yes." Some will respond by maintaining the above scenario is extreme and inappropriate. The fact is, however, that this scenario is simply a quantitative extension not qualitatively different from what has been occurring in the lowering of admission and course requirement standards that has been occurring for some time. So the question now becomes, not simply teaching effectiveness but teaching effectiveness at what level of learning, and by implication, academic standards. This is an issue that needs to be addressed nationally by faculty. Being well versed in logic as well as statistics, Scriven (1988), of course, understands this. In a similar context he notes,

> It's not even true that 'it all boils down to how much the students learn from the teacher': if it did, the teachers of mentally-retarded students would automatically be the worst teachers. In fact. they are often much better teachers than those teaching smart students, because smart students survive bad teaching better. (How many of the research studies naively treated "amount learned" as the criterion against which they "validated" the indicators?) p.7

And herein lies the ghost in the machine of most statistical validation studies of SEF---at their very best: There is nothing wrong with the statistics only with the meaning of what they are purportedly measuring. Thus the problem is not a flaw in the data or the measurement instrument, but a flaw in the measurer.

Finally, to conclude this section, the implications for SEF in general and for the issue of validity seem clear. The issue of validity of SEF, then, is not the primary issue it appears to be, and serves inadvertently to hide the significant issue of academic standards.[33] I will address this issue in relation to academic freedom and academic standards in more detail in my final paper. [34]

Conclusion

From most of the above cases---even given that, as challengers, the burden of proof has been on faculty --- it seems clear that the courts have not been kind to faculty with regard to student evaluations.[35] Some clearly see the courts various involvements in academic matters as detrimental to academic freedom. Arguably, rulings do often seem to shape it in inappropriate---and not so arguably---inconsistent and contradictory ways. "It is not clear, however," suggests Rebell (1990b), "that increased judicial involvement will have such a detrimental impact. In some measurement situations, courts have exhibited a sophisticated understanding of the complex judgmental factors at stake, and their insistence on thorough-going implementation of improved, fairer assessment devices has enhanced, rather than impeded, the development of professional standards" (p.340). He goes on to point out that, "because the state of the art concerning teacher-evaluation practices is at a sensitive developmental stage, extensive court intervention at this point can substantially influence---for better or worse---the future direction of basic practice in the field" (Rebell, 1990b, p.344). Thus whether increased judicial intervention in faculty matters will have a positive or a negative impact on professional evaluation practice depends on providing the courts with appropriate psychometric data and other scientific procedures.

Given the above rulings and the courts propensity to accept faculty/institutional agreements, it would seem as Kaplin and Lee advise, regarding academic freedom that "it is especially crucial for institutions to develop their own guidelines on academic freedom and to have internal systems for protecting academic freedom in accordance with institutional policy" (p. 192) would be especially true for a detailed SEF policy, especially including how the data is to be assessed.

The fourth and final paper will address the implications of court reasoning and rulings for academic freedom, standards, and instructional decisions.

Notes

1. Address correspondence to: Robert E. Haskell, Ph.D., Professor of Psychology, Department of Social and Behavioral Sciences, University of New England, Biddeford, Me. 04005. Email: haskellr@gwi.net. I would like to thank Professor John Damron, of Douglas College for continually

providing me with sources, support, and advice, and especially Professor William A. Kaplin, School of Law, Catholic University of America for his invaluable legal counsel and for reading a draft of this paper. Interpretive liberties with the legal material and any other problems and omitted legal nuances are my responsibilities.
[BACK to document]

2. As with my second paper (Haskell, 1997b), the focus here will be delimited to how the courts reviewed have addressed SEF issues within various legal challenges to the denial of academic freedom, tenure, promotion, and reappointment by institutions of higher education. There are multiple legal variables that define an action or influence an outcome in a particular case. Among them are the statutes or other sources of law being applied, the cause of action being asserted, the prescribed prima facie case, the allocation of burdens of proof, and the standards of judicial review (see, e.g., Kaplin and Lee, 1995, section 1.3 & section 1.4.3.6). For my purposes here, I will not be concerned with these variables. Accordingly, this paper will neither be concerned with the outcome of the legal rulings, nor with the complex legal reasoning on which the rulings were based. My purpose is to review the general reasoning of the courts on SEF from a "reasonable man" standard and from a policy point of view.
[BACK to document]

3. To the layman, legal rulings regarding SEF are a veritable thicket, often seeming that the use of context to differentiate one apparently similar case from another functions as a kind of ad hoc carte blanche to justify preconceptions and positions.
[BACK to document]

4. A largely neglected---or ignored---important function of education is its social function. Education is not just for the benefit of the individual but for the benefit of society. Like it or not, we in higher education have accepted the social function of certifying competence of our students entering into an increasingly complex world. The certifying function has become especially important since the introduction of vocational programs into university curricula.
[BACK to document]

5. In *Johnson v. University of Pittsburgh* (1977), the court said (7) "We have repeatedly approved the use of statistical proof where it reached proportions comparable to those in this case to establish a prima facie case of racial discrimination in jury selection cases . . Statistics are equally competent in proving employment discrimination. We caution only that statistics are not irrefutable. They come in an infinite variety and, like any other kind of evidence they may be rebutted. In short, their usefulness depends on all of the surrounding facts and circumstances" (8) The court further said in Footnote # 20: "Considerations such as small sample size may of course detract from the value of such evidence" " (p.1361).

In *Peters v. Middlebury College* (1977), it was maintained that (5) "A professor's value depends upon his creativity, his rapport with students and colleagues, his teaching ability, and numerous other intangible qualities which cannot be measured by objective standards" (p.860).

In *Fields V. Clark University* (1987), the court noted that (10)

Fields' "attacks" the university's use of her student evaluations because they were not gathered and evaluated according to accepted standards of scientific polling procedures. In response, the court agreed, saying, "She is probably correct. The use made of the student evaluations in her case, however, followed the practice at the defendant's university in other tenure decisions" (p.671).

In *Cynthia J. Fisher v. Vassar College* (1995), the court noted that (7) "statistical analyses may be a part of a plaintiff's effort to establish discriminatory treatment" (p.1209).

In *Yu Chuen Wei and the Vermont State Colleges Faculty Federation* (1995), the court ruled that (4) "The Court need not consider the accuracy of these administrative determinations, and that (24) tenure criteria "are not drawn with mathematical nicety." The board further ruled that (25) "the Dean and the President, both reviewed Grievant's student evaluations carefully. Their failure to take it a step further, and perform a statistical comparison of Grievant's student evaluations with those of other faculty members who have been granted tenure was not arbitrary and was reasonable; (26) Such a comparison is nowhere required by the Contract, [and] (27) we decline to hold such an involved comparison is necessary before a reasonable tenure determination can be made" (p.311).

In *Dr. Brian Maclean v. President of The University of British Columbia* (1991), the court concluded (38) "that the instrument was not perfect, that it had flaws, and that the very limited number of samples (because of the very limited number of courses and students surveyed over the period) impaired its reliability. (p.30). (39) "However, we accept the evidence of Dr. [X] that the instrument has some value, directed toward the specified factors. The court noted that (28) "One problem with the questionnaire is that it solicits bad points as well as good points. Despite that caveat, we conclude that the inclusion of the qualitative comments was not a significant error" (p.32).

In *Robert Kramer v. The President of the University of British Columbia* (1992), the Board said (19) Given certain Departmental procedures, "there is a danger that some negative class commentary will dominate the discussion and will not be the 'independent' opinion of all of the students. (20) This is especially true in the context of the direction to assess "effectiveness" versus "popularity" (p.10). They further noted, (18) Given that "There was no peer review at all; no member of the Department audited any of Dr. Kramer's lectures. There was, therefore, nothing to guide the Department but the student comments," and "no way to test the accuracy or fairness of the undoubtedly disturbing comments in Asian Studies" (p.10).

In *University of Regina Faculty Association v. University of Regina* (1993), The Board argued (6) that "the University was under an obligation to verify negative comments before acting on them" (p.4).

In *Christopher Turner v. The President of the University of British Columbia* (1993), the Board said, (7) "while not ignoring some student unhappiness with Dr. Turner's teaching style, we think that the comments and emphasis on the size of Dr. Turner's classes as evidence of poor teaching are open to objection and constitute errors of procedure and/or evidence" (p.6).

[BACK to document]

6. This is an important area but will not be dealt with here because student achievement scores as a measure of teaching effectiveness is almost exclusively used on the secondary level of education.
[BACK to document]

7. In *Dyson v. Lavery* (1976), the court found that despite questionable errors it concluded that administrative judgements were acceptable because, "they were sincere and grounded on some evidentiary basis" (p.111); and (5) "In the absence of a finding that same were sexually motivated, the administration's professional judgment must be respected" (p.111 all italics added).

In *William Sypher v. Vermont State Colleges Faculty Federation* (1982), (7) sufficient evidence exists from which the Dean and President could have reasonably concluded Sypher was not above average in his teaching effectiveness; (8) the Board went on to say that if they adopted the Colleges' view that Sypher was not reappointed because of his teaching effectiveness, no argument advanced by him defending his teaching was likely to persuade the President because his decision was made on the "vigor and variety of student criticisms" (p.135).

In *Carley v. Arizona Board of Regents* (1987), The court ruled (18) the University president was free to consider factual findings made by minority members of the academic freedom and tenure committee and any other evidence which he found relevant in determining whether to deny renewal of teaching contract to non tenured instructor. The president was not bound by factual findings made by majority members of committee (P.1103).

In *Yu Chuen Wei v. Vermont State Colleges Faculty Federation* (1995), it was noted that (28) The Dean and the President obviously had much experience in reviewing student evaluations, and could reasonably draw on that experience in each tenure review. (p.311); judgements "were not arbitrary or capricious and were exercised honestly upon due consideration,"....that Deans and Presidents have "much experience in reviewing student evaluations, and could reasonably draw on that experience" (p.311).

In *Dr. Brian Maclean v. President of The University of British Columbia* (1991), the court said, (40) The relevance and quality of the scores are "a matter of weight for the various decision-makers, and we assume that they were reasonably aware of the limitations of student evaluations and gave them the weight they deserve" (p.30).

In *Robert Kramer v. The President of the University of British Columbia* (1992), the board concluded, "In the final analysis, we feel that this review of the Head's comments on teaching, which would be the sole evidence upon which the Dean and the President could rely, shows that it was incomplete and might have been misleading" (p.12-14).

In *University of Regina Faculty Association v. University of Regina* (1993), he Board said teaching was wrongfully evaluated, but upheld denial of tenure on grounds of inadequate scholarship.

In *Christopher Turner v. The President of the University of British Columbia* (1993), The board concluded that (11) "there were sufficient errors of procedure and/or evidence to return the case for reconsideration" (p.11).
[BACK to document]

8. In *Lieberman v. Grant* (1979), Lieberman attempted to introduce approximately ten personnel files concerning the tenure proceedings of other faculty in the English department for comparison. (6) Recognizing that such evidence would have had some minimal probative value, the Court, exercised its discretion under Fed. R.Ev. 403, and excluded it on the ground that "such probative value would be substantially outweighed by the delay and waste of time, which introduction of such evidence would have necessarily entailed....The plaintiffs case without such evidence seemed almost interminable, consuming 52 trial days over a two-year period. That is long enough" (p.873).

In Fields V. Clark University (1987) notes but does not admonish the non separation of student remarks from small seminar courses and those from large lecture classes.

In *Cynthia J. Fisher v. Vassar College* (1995), the district court found (2) that the biology department distorted Fisher's teaching recommendations by (3) "selectively exclud[ing] favorable ratings," by "focus[ing] on the two courses in which Dr. Fisher had difficulties" and (4) by "applying different standards to her than were applied to other tenure candidates" (p.1209).

In *Yu Chuen Wei v. Vermont State Colleges Faculty Federation* (1995), it was noted that (19) "The statistical comparison demonstrates that Grievant was evaluated higher by students than her [male colleague] with respect to upper level classes, but that (20) [male colleague] was evaluated higher than Grievant in lower level classes. Given (21) this "mixed" result, the statistical comparison of evaluations does not demonstrate by a preponderance of the evidence that Grievant's students rated her the same, or better, than [male colleague]" (p.305). Wei maintained that (16) her students rated her the same or higher than the male colleague's students rated him. The Board disagreed, saying, (19) "We note that the comparison offered by Grievant is somewhat weak since [male colleague] was tenured in 1988, and those student evaluations of his which were compared with Grievant post-dated his tenure review by a number of years...further saying, "we decline to hold such an involved comparison is necessary before a reasonable tenure determination can be made" (p.305).

In *Dr. Brian Maclean v. President of The University of British Columbia* (1991), the Board noted that (19) the reviewing faculty held in-class discussions about his teaching.

In *Robert Kramer v. The President of the University of British Columbia* (1992), Kramer argued that the most significant mistake was the failure to consider all aspects of his teaching. For example, only his teaching in 1989-90 was considered, whereas (9) he had taught a wide range of courses over the previous three years (10) had three new courses that year, (11) plus a graduate course. Moreover, (17) The department head indicated that his teaching was not up to the departmental "standard." The standard appeared to be the performance of the tenure-track faculty, though Kramer was one of the most junior faculty members (p.8). (15) Only one of the more than thirty numerically rated questions was used: "Rate instructor bad to good." (16) While a number of negative student comments were quoted in the department Head's letter, there were a number of very positive comments, and these were not mentioned at all.

In *Christopher Turner v. The President of the University of British Columbia* (1993), the Dean said, "there were few students in undergraduate literature courses since 1986/7---(3,8, and 6 respectively," thus mistaking student 'response' figures for actual student enrolment. The Board concluded that (5) "This misunderstanding is in our opinion sufficient in itself for a reconsideration, since teaching was the focus..." (p.3), and (7) "we think that the comments and emphasis on the size of Dr. Turner's classes as evidence of poor teaching are open to objection and constitute errors of procedure and/or evidence" (p.6).

[BACK to document]

9. Given the extensive variation of rulings on SEF cases, from the perspective of a non legal professional it seem that legal reasoning carries the use of contextual analysis and variables to an extreme, making it possible---and justifiably legally---to rule just about anyway a court wants to rule. The logical extension of such reasoning would lead to each case being unique and nonsignificantly related to any other case.

[BACK to document]

10. In *Johnson v. University of Pittsburgh* (1977), the court noted that (10) "It has also been pointed out that in some cases difficult courses have to be given to the students and the material is such that it is difficult for even the best teacher to get it across.

In *Carley v. Arizona Board of Regents* (1987), he (7) characterized his professional style as being a "demanding teacher contrary to some student expectations," (8) Because of this, he maintained his popularity suffered and resulted in low student evaluations, (9) examination of his student comments indicated that Carley was correct in his assessment as 61% (49 out of 80) negative student comments focused on these values. The court ignored these findings.

In *Dr. Brian Maclean v. President of The University of British Columbia* (1991), it was noted that (21) While the knowledge, interest and enthusiasm of Dr. MacLean were acknowledged, "the problem appeared to be one of style or personality."

In *Robert Kramer v. The President of the University of British Columbia* (1992), the Board noted that (26) It was obvious that almost all of the classes were upset about an examination which was considered more geography than Asian Studies, and (27) they didn't like the marking. (28) They also felt the workload was far too heavy for an "introductory" course. The Board apparently only noted this variable.

[BACK to document]

11. In *Johnson v. University of Pittsburgh* (1977), the court said, "It is also obvious that the court and the administration of universities cannot permit students to exercise a veto over professors who may be world renowned scientists and yet if the students rate them unfavorably can be terminated at any time because of unpopularity" (p.1366-7).

In *Carley v. Arizona Board of Regents* (1987), he (8) he maintained his popularity suffered as reflected in his low student evaluations

In *Robert Kramer v. The President of the University of British Columbia* (1992), he maintained that (14) Student evaluations were considered from the standpoint of his popularity, not his effectiveness.

In *Brian Maclean v. President of The University of British Columbia* (1991), (35) The Faculty Agreement specified that "Evaluation of teaching shall be based on the effectiveness rather than the popularity of the instructor." Courts have ruled in various directions on this issue.

In *Robert Kramer v. The President of the University of British Columbia* (1992), the board noted (21) "As for the 'popularity vs. effectiveness' debate, a discouraging or hostile attitude is a part of effectiveness as much as it is of popularity" (p.8).

In *Christopher Turner v. The President of the University of British Columbia* (1993), the Board ruled, (8) while popularity is not competence nor effectiveness, to the extent that it encourages students it has some relation to both" (p.7).

[BACK to document]

12. There may well be research showing that being a popular teacher affects learning on elementary and secondary levels of education, I know of no such rigorous research on the post secondary level. In my view, one of the problems is that all too often we automatically transfer findings from elementary and secondary levels to higher education.

[BACK to document]

13. In *Johnson v. University of Pittsburgh* (1977), the court noted that it (5) "has placed little reliance on students' surveys....students in a given course rating a teacher, or professor, some of them as excellent, others as terrible and in between, many who say passable, mediocre etc.... we cannot say it was unreasonable for the tenured faculty to consider this along with other matters" (p.1359). (8) "It is also obvious that the court and the administration of universities cannot permit students to exercise a veto over professors who may be world renowned scientists" (p.1366-7). A similar view was expressed in *Yu Chuen Wei v. Vermont State Colleges Faculty Federation* (1995).

In *Peters v. Middlebury College* (1977), the court gave some weight to an administrative devaluing of a set of positive student evaluations of a faculty that said (2) "The department chair sent a letter to the president of the college, saying, " The course of action I recommend is not likely to be popular with students who, though they in part recognize her intellectual limitation, are warmly responsive to her enthusiasm, energy, openness and ready human concern" (p.860).

In *Carley v. Arizona Board of Regents* (1987), the court said, (23) "Carley has cited no authority that relying primarily or solely on student evaluations would be impermissible. We have found none" (p.1105, italics added).

In *Guam Federation of Teachers v. The University of Guam* (1990), the Guam Federation of Teachers challenged the use of SEF in tenure and promotion decisions (Blum, 1990). The Board (1) ruled to remove anonymous student evaluations from professors' tenure files, (2) The union said the use of SEF violated the union's contract with the university, (3) which provides that anonymous documents or those "based on hearsay" should not be included in a faculty member's file, (4) The court further ruled that (5) students should be made aware of the purpose and ramifications of their evaluations, and (6) anonymous student evaluations should not be used.

In *Robert Kramer v. The President of the University of British Columbia* (1992), the Board noted that (18) "The most important perceived error in the teaching evaluation, in the opinion of the Board, is the reliance solely upon the student evaluations and written comments for the 1989 course evaluations. There was no peer review at all; no member of the Department audited any of Dr. Kramer's lectures" (p.10).

In *University of Regina Faculty Association v. University of Regina* (1993) a Canadian Arbitration Board ruled that (3) "With respect to teaching, it is our opinion that the evidence of unsatisfactory performance is very weak indeed ...It is important to note that the basis of the comments, particularly the negative ones in the fall of 1992, were written student assessments... [and] Although these assessments are expressly recognized in Art. 17.19 of the collective agreement, to base important career decisions on them only does not seem justified" (p.4). The Board further ruled (4) that tenure decisions could not be based solely on assessments which were completed by students who had never been made aware of the ramifications of their statements. (5) [I]f evaluations are to be used for serious career development purposes those completing them should be aware of the potential consequences of their participation" (p.4) (8) "To base serious career decisions narrowly on student evaluations is not to be encourage... (9) If teaching is to be seriously evaluated for career purposes, whether for positive or negative purposes, it seems incumbent upon Faculties not to rely only on classroom administered evaluations but to broaden the base of assessment" (p.4).

In *Christopher Turner v. The President of the University of British Columbia* (1993), the Board ruled, (9) while the [Faculty Association] Agreement permits, but does not mandate either student reviews or peer reviews, and the methods of assessment 'may vary', we do conclude that the reliance placed on these very limited student reviews must have been great, since there was no other evaluation referred to. Where there is no other evidence sought, student comments will have an apparent importance and credibility that they may not deserve... (10) We would strongly recommend peer review in the reconsideration which we are requiring" (p.7). The board further noted that (8) "This board has been asked on a number of occasions to pass judgment on the relevance of student evaluations to the [Faculty Association] Agreement criteria for good teaching. Good teaching is an elusive concept. Students may not be good judges during a course; their judgment might be quite different several years later in life. (p.7).

[BACK to document]

14. In *Dyson v. Lavery* (1976), a student evaluation ranked her 46th of 48 teachers.

In *Lieberman v. Grant* (1979), the court noted (4) a compilation of student ratings showed that the cumulative ratings for members of the department ranged from a low of 4.09 to a high of 8.95. She had a cumulative rating of 7.06, which ranked her 12th out of the 15 junior faculty members. The 7.06 figure included the ratings from a previous semester in which the plaintiff received a rating of 8.18. Prior to this rating in the spring of 1972, the plaintiff's cumulative rating was 6.7.

In *Carley v. Arizona Board of Regents* (1987), it was noted that

(1) of the 13 faculty in his department of art, he was ranked fifth, (2) by his chairman he was ranked 7th, (3) student evaluations, however, ranked him last: 13th of 13 (p.1105).

In *Robert Kramer v. The President of the University of British Columbia* (1992), the court noted (24) scores in the other two courses were higher---3.45 in one, 3.91 in another, against a "faculty average" of 4.22. The board further noted, "In the result, one got a 2.82 and one got a 3.07...the difference is statistically invalid in any event" (p.10).

[BACK to document]

15. In *Dyson v. Lavery* (1976), the course said (1) "A number of students apparently had voiced displeasure over the quality of her class preparation and presentation" (p. 111 (3) "These *impressions*" said the court, "were largely confirmed after the initial decision to not rehire her had been made, by a student evaluation that ranked her 46th of 48 teachers in the Business Department" (p.111, italics added).

In *Johnson v. University of Pittsburgh* (1977), the court said, (3) "we have *the instance* referred to in Finding 27 (p.1359, italics added).

In *Lieberman v. Grant* (1979), the court noted (3) based on complaints received from "*several students*," to the effect that Lieberman's interest in feminism caused her to ignore other themes in literature (p.873, italics added).

In *William Sypher v. Vermont State Colleges Faculty Federation* (1982), (1) some of the student comments noted that, "When students try to disagree he shoots you down and tries to degrade you in front of the class," (p.115), while others said, "encourages student participation as much as possible... encourages student to express their ideas freely and not worrying how 'dumb' it may sound...always wants you point of view." (P.115) (2) With regard to the numerical ratings, the Board's opinion was that (3) "regardless of a strong majority of students' rating his teaching as above average, (4) the existence of a significant minority of students feeling degraded, humiliated, and embarrassed can reasonably lead an evaluator to question a teacher's effectiveness" (p.115).

In *Yu Chuen Wei v. Vermont State Colleges Faculty Federation* (1995), the Board said, (22) "the statistical comparison does not take account of the comments made by students on the evaluation forms. Grievant's student evaluations are striking in how often mention is made of Grievant's communication difficulties, particularly language difficulties (p.304-5). The board further noted with respect to comments that while some students had written that she was a "slant eyed bitch," and that she should "go back to China," (30) "We also are not persuaded that the racism evident in the student evaluations of Grievant made student evaluation results unreliable. The percentage of evaluations in which racism by students was evident was approximately one percent of the total evaluations" (p.306).

In *Robert Kramer v. The President of the University of British Columbia* (1992), (2). The department Head viewed Kramer's 1989-90 course evaluations "with some alarm"....(4) Even more disturbing to the department Head was that a considerable number of students in their written comments stated that Dr. Kramer was biased, sarcastic, and hostile to the material and that a number of students had stated that Dr. Kramer's teaching would cause them to stay away from the Asian Studies department. (5) There were also some diametrically apposed positive comments" (p.10).

In *University of Regina Faculty Association v. University of Regina* (1993), The Board argued (6) that the University was under an obligation to verify negative comments before acting on them. Consequently, (7) the fact that Dr Jalan had received some negative evaluations from students could not be used to undermine the otherwise generally favorable comments he had received in his annual performance reviews" (p.4).

In *Dr. Brian Maclean v. President of The University of British Columbia* (1991), the court noted that (25) "With respect to the "qualitative" scores---i.e., the "comments," there was a clear error. The qualitative comments from a number of courses were read and commented on, and conclusions were drawn from them which went into the "file." Both Reviewing faculty read and commented on them, as did the Department Chair in her letter to the Dean. Yet the Dean had clearly stated in a departmental memo that the qualitative comments were not to be used for administrative or promotion purposes. (26) While in the abstract there is no reason why such comments would not be relevant, if the Department had a rule against their use, or in other words if they were "for the professor's eyes only," then it was a significant breach of Departmental rules to use them" (p.31). (27) In the opinion of the Board, so long as the comments were fairly presented, they offered the PAT [Promotion and Tenure Committee] and others a better balanced view of the teaching qualities and problems of Dr. MacLean than the quantitative statements alone" (p.31). (28) The court noted that "One problem with the questionnaire is that it solicits bad points as well as good points. Despite that caveat, we conclude that the inclusion of the qualitative comments was not a significant error" (p.32).

[BACK to document]

16. In *Johnson v. University of Pittsburgh* (1977), the court noted (2) they "approached this question of teaching ability with considerable doubt, in view of the fact that in prior years there does not appear to have been any criticism of her teaching and also in view of the fact that...there was evidence that the department chairman, had informed her after one of her lectures in 1971 what a great lecture it had been;" On the other hand, the court said (3) "we have the *instance* referred to in Finding 27 (p.1359, italics added).

In *Fields V. Clark University* (1987, it was observed (3) a few of which, from students in Fields' seminars, were "wildly enthusiastic" about her enthusiasm, commitment and presentations; (4) a few were ambivalent; (5) with a considerable number being extremely negative, particularly (6) with regard to her large lecture classes in basic courses in sociology.

In *Yu Chuen Wei v. Vermont State Colleges Faculty Federation* (1995), moreover, they said, (19) "The statistical comparison demonstrates that Grievant was evaluated higher by students than [her male colleague] with respect to upper level classes, but that (20) [the male colleague] was evaluated higher than Grievant in lower level classes. Given (21) this "mixed" result, the statistical comparison of evaluations does not demonstrate by a preponderance of the evidence that Grievant's students rated her the same, or better, than [male colleague]" (p.305).

In *Dr. Brian Maclean v. President of The University of British Columbia* (1991), it was noted that (20) In general, the in-class peer reports were mixed but favourable. The in-class discussions were more problematic. (p.30). (21) While the knowledge, interest and enthusiasm of Dr. MacLean were acknowledged, "the problem appeared to be one of style or personality." It was further noted that (29) "As against the low figures, they disclosed a number of good qualities in Dr. MacLean---enthusiasm for his subject, wide knowledge of the literature, much out of class assistance to students, and a commitment to seeking good work from students. (p.31). (30) The reviewing faculty report noted the comments about Dr. MacLean's "derogatory manner, biased opinion, unwillingness to listen," were matched by "clear, stimulating, very helpful after class." And, (31) "some students have told us that the comments made were not representative of the class as a whole and were unduly influenced by the process" (p.41). (32) "A number of students, both from earlier years and from his current classes, furnished letters of support, and in preparation for the appeal, some furnished affidavits with respect to particular matters such as the 'intimidation' discussion in Soc. 250 and events in Soc. 490 and 520 in the fall of 1989." (p.33)

In *Robert Kramer v. The President of the University of British Columbia* (1992). (16) While a number of negative student comments were quoted in the department Head's letter, there were a number of very positive comments, and these were not mentioned at all. (25) "We have examined all of these written comments. There was a very wide range of comments. There were not 29 comments saying sarcastic and biased comments; but there were certainly 29 comments which included either cynical, sarcastic, biased, insulting, negative, condescending, belittling, opinionated, arrogant, nihilist, and destructive.... (29) However, it would only be fair to add that there were a number of comments in favour of Dr. Kramer, stating that the student "liked the course immensely," "now interested in Asian Studies;" "helps create a relaxed atmosphere," "really enjoyed him," "very approachable and knowledgeable," "very enthusiastic," "captivates audiences with his humour," "very effective" (p.12). (30) "In the other two courses, both small, both Japanese language, there were also some negative comments" (p.12).

In *Christopher Turner v. The President of the University of British Columbia* (1993), the board noted that (6) "While there is no question of Dr. Turner's competence as a teacher at all levels, teaching evaluations for the last several years show that his effectiveness is marred by what students perceive as excessive formality, lack of enthusiasm and dullness....In a previous promotion attempt, his teaching was briefly described as "very competent" but student evaluations indicate further improvement to be "better than adequate" (p.2)

[BACK to document]

17. I wish to thank to Patrick B. Shaw, Attorney for AAUP for referring me to Ms. Linda Lott, Administrative Coordinator, Hofstra Univeristy Chapter, AAUP, who conducted a search for me of a faculty collective bargaining contract database being developed there. Ms. Lott searched the database with "several key words that relate to academic freedom, teaching methodology and student evaluations. The only word

that was identified in some of the contract provisions was 'student evaluation'"(Personal communication, March 21, 1997). It should be noted that very few explicit references in the contracts to the use of signed/unsigned SEF or the use/nonuse of comments were found in this developing database. Some of the instances found are:

At Rider University, the agreement stated "The College may not use course evaluations for purposes of discipline, promotion, or tenure, unless introduced for such purposes by the faculty member."

At Western Michigan University, the agreement stated "Only the ratings shall be included in all promotion, reappointment, merit, and tenure recommendations, together with such other evaluations of teaching competence as may be employed by faculty members and made available. Western agrees to consider all the evidence of teaching competence that is presented in evaluating teaching faculty and shall not use unsubstantiated structured comments in personnel decisions." I have already noted the ruling at the University of Guam (Blum D. E. (1990, October 3). which stated that (1) students not being made aware of the purpose and ramifications of their evaluations, (2) the anonymous nature of student evaluations, (3) the invalid analysis of SEF, and therefore, (4) SEF in effect being anecdotal and hearsay data. Since most SEF results are prepared anonymously, an instructor has no recourse to confront his/her evaluators. As will be addressed below, the anonymous nature of SEF is beginning to also be questioned by arbitration boards.

I am informed from a colleague at St. John's University (New York) that, though SEF are mandated, they are not used administratively. I suspect there are many more schools (likely those who have union contracts) that do not use SEF administratively or who limit its use. I might note here for those who maintain that without SEF used administrative that there is no quality control over instruction and that therefore student learning will suffer, to check with the schools who do not use SEF administratively for a reality check on their assumption.

[BACK to document]

18. In *Johnson v. University of Pittsburgh* (1977), the court said, "It is also obvious that the court and the administration of universities cannot permit students to exercise a veto over professors who may be world renowned scientists" (p.1366-7), noting, "It is obvious that a professor may be possessed of excellent qualifications as a research scientist and not necessarily be able to prove his or her worth as a teacher, concluding that, (9) "in cases where one has an outstanding scientist of national or international reputation, one may decide to promote and give tenure notwithstanding inability to come across as a teacher, this however is not one of those cases" (p.1366-7).

In *Yu Chuen Wei v. Vermont State Colleges Faculty Federation* (1995), (31) Wei's last claim charged that the College violated the Contract by denying her a promotion, even though both her scholarly performance and professional activities were exceptional. Article 22(E) of the College provides for otherwise granting promotion if the President decides that "performance in one of three areas has been exceptional" (p.314). The Board concluded that "Although Grievant had a significant publication record, most of it was developed before coming to Castleton" (p.315). (33) In terms of exceptional scholarship, Dr. Wei maintained she had solved a significant mathematical problem (apparently published). The Board's response was, (34) "although Grievant claimed to have solved the Erdos conjecture, [the]Dean reasonably concluded that she had not established that she actually had solved the conjecture. Under these circumstances, and given our consideration of the discrimination issue previously discussed, we conclude that (35) Grievant has not established discrimination. The Colleges reasonably, and based on legitimate reasons, concluded that Grievant had met the tenure standards in this performance area but that her performance was not exceptional" (p.315).

In *Dr. Brian Maclean v. President of The University of British Columbia* (1991), the court said, (34) "while a superior research and publication record cannot overcome a poor teaching record, it might tip the scales where the teaching record was 'on the edge'" (p.10).

[BACK to document]

19. I would like to again acknowledge the invaluable assistance of Professor William Kaplin (Personal conversation, May 28, 1997). In constructing this section on the generalization of the principles disparate treatment and impact, I have gone where wiser and more skilled sailors would perhaps have elected not to sail. Without Bill's counsel I would clearly have sailed off the edge of this legal world. As it is, I may be dangerously close. But since my intent in not a strictly legal one I can perhaps be given some latitude.

[BACK to document]

20. Statistically, a case can be made that the racist comments of one percent represent a larger (but unknown) number of racist attitudes that simply were not overtly stated. Thus even one percent overtly racist comments should be a red-flag to look deeper into such a situation. And given a commitment to eradicating racial and gender discrimination, should not the overt one percent (plus) evidence of racism on a SEF be treated more seriously?

[BACK to document]

21. For example, in *Fields V. Clark University* (1987), the court noted that (10) Fields' "attacks" the university's use of her student evaluations because they were not gathered and evaluated according to accepted standards of scientific polling procedures. In response, the court agreed, saying, "She is probably correct. The use made of the student evaluations in her case, however, followed the practice at the defendant's university in other tenure decisions" (p.671).

[BACK to document]

22. Given the current attitude of the courts in non EEOC cases toward accepting the institutional interpretation of the SEF data, for those covered as members of a protected group ---assuming sufficient *prima facie* evidence to do so, and possible illegalities notwithstanding---using one's protected status could be used as a strategy for insuring a rigorous analysis of one's SEF data.

[BACK to document]

23. Once again, numerous legal variables prohibit such generalization. See Kaplin and Lee, 1995:

> ***1.4.3.6.*** *Standards of judicial review and burdens of proof.*

Postsecondary institutions have numerous processes for making internal decisions regarding the status of faculty, students, and staff, and for internally resolving disputes among members of the campus community. Whenever a disappointed party seeks judicial review of an institution's internal decision, the reviewing court must determine what "standard of review" it will apply in deciding the case. This standard of review establishes the degree of scrutiny the court will give to the institution's decision, the reasons behind it, and the evidence supporting it. Put another way, the standard of review helps establish the extent to which the court will defer to the institution's decision and the value and fact judgments undergirding it. The more deference the court is willing to accord the decision, the less scrutiny it will give to the decision and the greater is the likelihood the court will uphold it. Issues regarding standards of review are thus crucial in most litigation.

In turn, standards of review are related to the "burdens of proof" for the litigation. After a court determines which party is responsible for demonstrating that the institution's decision does or does not meet the standard of review, the court allocates the burden of proof to that party. This burden can shift during the course of the litigation (see, for example, Section 3.3.2.1). Burdens of proof also elucidate the elements or type of proof each party must submit to meet its burden on each claim.

1.4.3.6. *Standards of Judicial Review and Burdens of Proof 35* or defense presented. Such issues are also critical to the outcome of litigation and can become very complicated (see, for example, Section 3.3.2.1).

There are many possible standards of review (and likewise many variations of burdens of proof). The standard that applies in any particular litigation will depend on numerous factors: the type of institution subject to the review (whether public or private); the type of claim that the plaintiff makes; the institution's internal rules for reviewing decisions of the type being challenged; the character of the contractual relationship between the institution and the party seeking court review; and the common law and statutory administrative law of the particular state (see this volume, Section 1.3.1), insofar as it prescribes standards of review for particular situations. At a subtler level, the court's selection of a standard of review may also depend on comparative competence-the court's sense of its own competence, compared with that of the institution, to explore and resolve the types of issues presented by the case.

If a court is reviewing the **substance of a decision (whether the institution is right or wrong on the merits)**, it may be more deferential than it would be if it were reviewing the adequacy of the procedures the institution followed in making its decision-the difference being attributable to the court's expertise regarding procedural matters and relative lack of expertise regarding substantive judgments (such as whether a faculty member's credentials are sufficient to warrant a grant of tenure).

[BACK to document]

24. Again I am indebted to Bill Kaplin for this important point.

[BACK to document]

25. An assumption that students are qualified evaluators of teaching effectiveness I will deal with in the following paper.

[BACK to document]

26. I am not against using SEF for feedback to faculty. Some method of assessing faculty teaching effectiveness needs to be developed. No profession can be completely self policing. In terms of using SEF to assess teaching effectiveness and its use in tenure, promotion, reappointment, and merit salary increases, however, we need to proceed much more carefully than we have. As an initial general resolution to the problem of their use, I suggest that it be used as a "red flag" that can then set in motion a systematic faculty inquiry into the situation.
[BACK to document]

27. One teaches in such a classic mode today at one's litigious peril. See Pinsker (1989).
[BACK to document]

28. Scriven further suggests that "Unfortunately, most of what you see in a classroom are the features of teaching style, and you can't use any of them, because no amount of research can justify you in counting off brownie points for style against demonstrably badly or well-performed duties. You might as well try subtracting 10% of the purse from a professional golfer's pay-out on the grounds that he or she has an inelegant swing. You must get data on all duties; and when you have that, why would you need anything else?" (p.6).
[BACK to document]

29. SEF has become politicized. At a recent faculty development meeting on campus a well known "consultant" from a business school who has published books and articles of faculty evaluation was brought in by administration to work with faculty on a SEF and teaching portfolios. I mentioned Scriven's work to him. The consultants flip response was to call Scriven's views "fringe." I let it be known that this kind of non scholarly and ad hominem response was not acceptable. He has been critiquing and developing methods of assessment for some time. He takes a rigorous no-nonsense, yet practical approach to evaluation. For readers who may not know of Scriven's background his work, he is a philosopher of science of international reputation who was co-editor of the foundational series, *The Minnesota Studies in the Philosophy of Science*. In 1967 he coined the terms *formative* and *summative* evaluations that are coin of the realm today in evaluation research (Scriven 1981); he founded and was the first president of what is now the American Evaluation Association. Thus to call Scriven's views "fringe" is at best the epitome of arrogance, at worse, its more likely due to ignorance.
[BACK to document]

30. Most rigorous statisticians, however, would as a matter of course agree with Scriven. One of the first things we learn in psychological research is that statistical correlations can not be automatically applied to individuals. How this is generally transferred even by those who understand it in one domain to a different domain is quite another---yet important educational---question in itself.
[BACK to document]

31. Research into transfer of learning suggests many traditional methods that appear to produce effective learning may in fact be counterproductive

relative to long term transfer of learning. For example, it's generally the case that immediate feedback during learning results in more efficient learning. It therefore seems to logically follow that immediate feedback during learning would result in more efficient transfer of learning. Recent findings, however, indicate that under certain conditions *delayed* feedback is more efficient. Other examples point out the difference between understanding typical learning principles on the one hand, and principles of transfer of learning (i.e., generalization of learning and long term application) on the other. Schmidt and Bjork (1992), note "we have repeatedly encountered research findings that seem to violate some basic assumptions about how to optimize learning in real-world settings." For example, increasing the frequency of information about errors to learners during practice improves their performance. The fact is, that increasing the frequency about errors can work in just the opposite manner for long term retention and for transfer. Further counterintuitive effects come from research showing that increasing the variability of a task during practice depresses performance during training, but may increase transfer of performance after training when conditions are altered from the original training situation. Still other data show that performance on solving a puzzle is virtually perfect with no delay between instruction and application, but rapidly declines as the delay is increased (e.g., where periods of delay are two weeks and one, two, three, and four months). In contrast, performance on a similar puzzle was worse than performance on the same puzzle at first, but stayed relatively constant over a delay of four-months. In other words, the transfer effect was much more persistent than the specific effects of learning a particular puzzle. Singley and Anderson (1989). And so it goes.
[BACK to document]

32. As I was writing this section (5/10/97), I heard on the NBC national news about Jon Westling, (1995) the new President and former Provost of Boston University in trouble for suggesting that disability laws requiring special standards for students are contributing to lowering academic standards. I then went on the internet and found the following: "The disability laws are sacred cows, but they must at the very least be tethered so that they cannot be used to force universities to lower academic and other standards."
[BACK to document]

33. This section should avert the frequent "he-said-she said, you-show-me-yours-I'll-show-you-mine" approach to assessment of the SEF validity literature by administrators and some faculty, where the apparent conflicts in the literature are rationalized away by simply saying "well, some studies show that SEF are largely not valid, but other studies show it is valid."
[BACK to document]

34. In my first paper (Haskell, 1997a), I suggested that one of the reasons that SEF has not been viewed as a threat to academic freedom and generated more interest is because many do not consider it high status research and do not see its encompassing implications for quality education---even after reading some of the findings on SEF. Many of my colleagues, including those who have basically supported my efforts and

work on the issue have said to me "Well, you've made your point about SEF by making us aware of it and by publishing articles, why don't you now put it aside and get back to your *real* scholarship."
[BACK to document]

35. As one scholar (Damron, personal communication, April, 1997) who read a draft of this paper observed: from "the legal decisions you review in your paper it is clear that untenured and/or politically incorrect faculty are often considered to be "fair game" by administrators, with literally _any_ superficially plausibly excuse serving as a rationale for dismissal. Use of such strategies reveal that faculty are often regarded as little more than term employees who are as disposable (and replaceable) as tissues. Clearly, there is a very serious ethical issue here, and a hugely hostile attitude toward academic freedom and faculty in general....the great variety of decisions you've reviewed and their assorted implications for the coherence and ethics of the legal processes that gave rise to them...it seems to me that many judges and arbitration panelists have little sense of how to proceed in hearing[s] involving academics."
[BACK to document]

References

Abrami, P.C., Dickens, W.J., Perry, R.P., & Leventhal, L. (1980). Do teacher standards for assigning grades affect student evaluations of instruction? *Journal of Educational Psychology*, *72*, 107-118.

Abrami, P.C. (1989). How should we use student ratings to evaluate teaching? *Research in Higher Education*, *30*(2). 221-27.

Anderson, U., & Schadewald, M., (1991). Analogical transfer and expertise in legal reasoning. *Organizational Behavior and Human Decision Processes*, *48*, 272-290.

Baez, B., & Centra, J. (1995). Tenure, promotion, and reappointment: Legal and administrative implications. *Ashe-eric Higher Education Report No. 1*.

Barnett, L.D. (1996). Are teaching evaluation questionnaires valid? Assessing the evidence. *Journal of Collective Negotiations in the Public Sector*, *25*(4). 335-349.

Bauer, H.H. (1996). The New Generations: Students who don't study (prepared for the symposium, "The Technological Society at Risk" Annual Meeting, AOAC International, Orlando (FL), 10 September 1996). Email hhbauer@vt.edu for a copy of the article.

Blum, D.E. (1990, October 3). U. of Guam removes evaluations from files. *The Chronicle of Higher Education*, Section: Personal & Professional, p. A21

Brian Maclean v. President of The University of British Columbia (1991). Held at Vancouver, B.C. January 28--June 20.

Brimelow, Peter (1996). Devalued diplomas. *Forbes*, April 22.

Carley V. Arizona Board of Regents, 737 P.2d 1099 (Ariz. App. 1987).

Cashin, W.E. (1995). Student ratings of teaching: The research revisited. *IDEA Paper No. 32*. Manhattan, KS: Kansas State University, Center for Faculty Evaluation and Development.

Chacko, T.I. (1983). Student ratings of instruction: A function of grading standards. *Educational Research Quarterly*, *8*(2), 19-25.

Cohen, P.A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research*, *51*, 281-309.

Cynthia J. Fisher v. Vassar College (1995). United States Court of Appeals for the Second Circuit Nos. 1179, 1303, 2275 Docket Nos. 94-7737, 94-7785, 94-9125

Christopher Turner v. The President of the University of British Columbia (1993). Arbitration Appeals Board Case No.: AI-256 Vancouver, B.C. July 19,20 and 21.

Cohen, P.A. (1983). Comment on 'A Selective review of the validity of student ratings of instruction.' *Journal of Higher Education*, *54*, (no) 4.

Copeland, J.D., & Murry, J.W., Jr. (1996). Getting tossed from the Ivory Tower. *Missouri Law Review*, *61*, 233-327.

Divoky, J.J., & Rothermel, A. (1988). Student perception of the relative importance of dimensions of teaching performance across type of class. *Educational Research Quarterly*, *12*, 40-45

Dowell, D.A., & Neal, J.A. (1982). A selective view of the validity of student ratings of teaching. *Journal of Higher Education*, *53*, 51-62.

Dyson V. Lavery, 417 F.supp. 103 (E.d. Va. 1976).

Dawes, R.M., Faust, D., & Meehl, P.E. (1989). Clinical versus actuarial judgment. *Science*, *243*, 1668-1674

Dilts, D.A., Samavati, H., Moghadam, M.R., & Haber, L.J. (1994). Student evaluation of instruction: Objective evidence and decision making. *Journal of Individual Employment Rights*, *2*, 73-86.

Ellul J. (1964). *The technological society*. New York: Vintage Books.

Faust, D., Guilmette, T.J., Hart, K., Arkes, H.R., Fishburne, F.J., & Davey, L. (1988). Neuropsychologists' training, experience, and judgement accuracy. *Archives of Clinical Neuropsychology*, *3*, 145-163.

Feldman, K.A. (1983). Seniority and experience of college teachers as related to evaluations they receive from students. *Research in Higher Education*, *18*, 3-124

Fields V. Clark University (1986). Civil Action No. 80-1011-s, United States District Court for the District of Massachusetts, 40 Fair Empl. Prac. Cas. (Bna) 670, March 14,; Vacated and Remanded May 8, 1987.

Fighting grade inflation. (1994) *Science*, *264* (27 May), 1255.

Franklin, J., & Theall, M. (1990). Communicating student ratings to decision makers: Design for good practice. In Theall, M. & Franklin J. (Eds.), *Student Ratings of Instruction: Issues For Improving Practice*. San Francisco: Jossey-Bass.

Garb, H.N. (1989). Clinical judgment, clinical training, and professional experience. *Psychological Bulletin*, *105*, 387-392.

Golding, M.P. (1984). *Legal reasoning*. New York: Alfred A. Knopf.

Greenwald, A.G. (1996). *Applying social psychology to reveal a major (but correctable) flaw in student evaluations of teaching*. University of Washington, Draft Manuscript, March 1, submitted for publication.

Greenwald, A., & Gillmore, G. (1997). No pain, no gain? The importance of measuring course workload in student ratings of instructions. *Journal of Educational Psychology* (forthcoming).

Haskell, R.E. (1997a). Academic freedom, tenure, and student evaluation of faculty: Galloping polls in the 21st century. *Education Policy Analysis Archives*, *5*(6) [Refereed Online Journal]. Available: http://olam.ed.asu.edu/epaa/v5n6.html

Haskell, R.E. (1997b). Abridgement of academic freedom, promotion, reappointment and tenure rights by the administrative use of student evaluation of faculty: (Part II) Views from the court. *Education Policy Analysis Archives*, *5*(6), [Refereed Online Journal]. Available: http://olam.ed.asu.edu/epaa

Hayes, S.C. (1991). The emperor's clothes: Examining the 'delusions' of professional psychology: The healthy skepticism of David Faust. *Science*, *1*, 22-25.

Holmes, D.S. (1972). Effects of grades and disconfirmed grade expectancies on students' evaluations of their instructor. *Journal of*

Educational Psychology, *63*, 130-133.

Howard, G.S., & Maxwell, S.E. (1982). Do grades contaminate student evaluations of instruction? *Research in Higher Education*, *16*, 175-188.

Howard, G.S., & Maxwell, S.E. (1980). Correlation between student satisfaction and grades: A case of mistaken causation? *Journal of Educational Psychology*, *72*, 810-820.

Howard, G.S., Conway, C.G., & Maxwell, S.E. (1985). Construct validity of measures of college teaching effectiveness. *Journal of Educational Psychology*, *77*, 187-196.

Johnson V. University of Pittsburgh (1977). 435 F.supp. 1328 (W.d. Pa.).

Kaplin, W.A., & Lee, B. (1995). *The law of higher education: A comprehensive guide to legal implications of administrative decision making*. San Francisco: Jossey Bass.

Larkin, J., McDermott, J., Simon, & Simon, H. (1980). Expert and novice performance in solving physics problems. *Science*, *208*, 1335-1342.

Lasch, C. (1979). Schooling and the new illiteracy (Chapter Six), in *The culture of narcissism: American life in an age of diminishing expectations*. New York: Warner Books.

Lemons, J. (ed.) (1996). *Scientific uncertainty and environmental problem-solving*. Cambridge, MA: Blackwell Science

Lemons, J., Shrader-Frechette K., & Cranor C. (in press). *The precautionary principle: Scientific uncertainty and type-I and type-II errors*. Foundations of Science.

Leo, J. (1996). No books, please; We're students. *U.S. News and World Report*, September 16.

Levi, E.H. (1949). *An introduction to legal reasoning*. Chicago: University of Chicago Press.

Lieberman V. Grant (1979). 474 F.supp. 848 (D.conn.).

Lovelace v. Southeastern Massachusetts University (1986). 793 F.2d 419 (1st Cir.).

Marchant, G., Robinson, J., Anderson, U., & Schadewald, M. (1991). Analogical transfer and expertise in legal reasoning. *Organizational Behavior and Human Decision Processes*, *48*, 272-290.

Marsh, H.W., & Dunkin, M.J. (1992). Students' evaluations of university teaching: A multidimensional perspective. *Higher Education: Handbook of Theory and Research*, *8*, 143-233.

Marsh, H.W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, *76*, 707-754.

Marsh, H.W. (1982). Validity of students' evaluations of college teaching: A multitrait-multimethod analysis. *Journal of Educational Psychology*, *74*, 264-279.

Marsh, H.W. (1980). The influence of student, course, and instructor characteristics on evaluations of university teaching. *American Educational Research Journal*, *17*, 219-237.

McKeachie, W.J. (1979). Student ratings of faculty: A reprise. *Academe*, *65*, 384-397.

Neath, I. (1996). How to improve your teaching evaluations without improving your teaching. *Psychological Reports*, *78*, 1363-1372.

Peters V Middlebury College (1977). 409 F.supp. 857 (D. Vt.).

Pinsker, S. (1989). Teaching in a litigious age. *Change*, *21*, (July/August), 50-54.

Powell, R.W. (1977). Grades, learning, and student evaluation of instruction. *Research in Higher Education*, *7*, 193-205.

Rabinowitz, J. (1993). Diagnostic reasoning and reliability: A review of the literature and a model of decision-making. *The Journal of Mind and Behavior*, *14*, 297-316

Rebell, M.A. (1990). Legal issues concerning teacher evaluation. In Millman, J., & Darling-Hammond, L., (Eds.), (pp. 337-355) *The new handbook of teacher evaluation*. Beverly Hills, CA: SAGE Publications.

Robert Kramer v. The President of the University of British Columbia (1992). Arbitration Appeals Board Case No. AI-245. Held at the University April 22 and 23.

Sacks, Peter Generation X Goes to College, Chicago & LaSalle (IL): Open Court, 1996.

Singley, M.K., & Anderson, J.R. (1989). *The transfer of cognitive skill*. Cambridge, MA: Harvard University Press (p.199).

Scheelhaase v. Woodbury Central Community School District (1973), In Rebell, M, A. (1990). Legal issues concerning teacher evaluation. In Millman, Jason& Darling-Hammond, Linda, (Eds.), (pp. 337-355) The new handbook of teacher evaluation. Beverly Hil ls: CA. Sage Pub.

Schmidt, R. A., & Bjork, R.A.(1992). New conceptualizations of practice:

Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*. 207-217.

Scriven, M. (1988). The new crisis in teacher evaluation: The improper use of research-based' indicators. *Professional Personnel Evaluation News*. (January) p.4.

Scriven, M. (1987). Validity in personnel evaluation. *Journal of Personnel Evaluation in Education, 1*, 923.

Scriven, M. (1991). *Evaluation Thesaurus*. (4th Ed.) Newbury Park, CA: Sage Publications

Scriven, M. (1993). The validity of student ratings: In Teacher evaluation. *Evaluation & Development Group*.

Scriven, M. (1995). A unified theory approach to teacher evaluation. *Studies in Educational Evaluation, 21*, 111-129

Scriven, M. (1981). Summative Teacher Evaluation. In J. Millman (Ed.) *Handbook of teacher evaluation* (pp. 244271). National Council on Measurement in Education. Beverly Hills, CS: Sage Publications.

Simon, W.E. (1996). The dumbing down of higher education. *Wall Street Journal*, March 19.

Snyder, C.R., & Clair, M. (1976). Effects of expected and obtained grades on teacher evaluation and attribution of performance. *Journal of Educational Psychology, 68*, 75-82.

Stern, J., & Flynn, P.D. (1995). Students propose a course of action for grade inflation. *The Bucknellian*. (Feb, 20).[Online] Available:

Sunstein, C.R. (1993). On analogical reasoning. *Harvard Law Review, 106*, 741-79.

University of Regina V. University of Regina Faculty Association and Dr. Pradeep Jalan. (1993). Case No.: AI-298 Regina, Saskatchewan, July 9.

Vasta, R., & Sarmiento, R.F. (1979). Liberal grading improves evaluations but not performance. *Journal of Educational Psychology, 71*, 207-211.

Westling, J. (1995). Getting Government Out of Higher Education. Boston University, delivered May 3. Available Online: http://www.heritage.org/heritage/library/categories/education/lect533.html

Worthington, A.G., & Wong, P.T.P. (1979). Effects of earned and assigned grades on student evaluations of an instructor. *Journal of Educational Psychology, 71*, 764-775.

William Sypher v. Vermont State Colleges Faculty Federation. (1982). 5

VLRB 102.

Yu Chuen Wei and the Vermont State Colleges Faculty Federation. (1995) 18 VLRB 261.

Appendix: A Non Litigated Case of SEF Used in the Denial of Tenure and Reappointment

The following is a case of the denial of tenure and reappointment primarily on grounds of apparent non outstanding teaching as measured by SEF. The case illustrates most of the issues discussed in this paper. As in most such denials, Dr. Tichenor did not go to litigation as litigation is costly given the low odds of the courts finding in favor of faculty. The case is an example of the likely thousands of such cases (given that there are over 3,500 colleges in the U.S.) that do not go to litigation. In most respects this case is probably typical relative to the inappropriate use of SEF in faculty dismissals.

The case involves Dr. Linda L. Tichenor, assistant professor of biology, who was denied reappointment during the 1996/97 academic year at a small private, student tuition-dependent university. The SEF material presented here was given to me for use here by Dr. Tichenor. It is from a draft paper by her that will appear in a special monograph series to be published by *The Society of College Science Teaching*, edited by Mario W. Caprio.

The unique aspect of this case of non reappointment due to SEF, ostensibly reflecting teaching ineffectiveness, is that Dr. Tichenor graduated from Idaho State University, Doctorate of Arts program in biology. The program, established by the Carnegie-Mellon Foundation in the late 1960's, specifically prepares future faculty for college science teaching pedagogy as well as for teaching a broad range of courses within the discipline of life sciences. The program requires a breadth of background in the biological sciences, knowledge of learning processes and pedagogy, awareness of the objectives of an undergraduate education, and the development of a sound educational philosophy. "As a result of my training," says Dr. Tichenor, "my classroom was to become my 'research bench.'" This is fairly unique in higher education. Few higher education faculty have any formal training in teaching (with the exception of faculty in the discipline of Education). Granted, while a degree in teaching does not guarantee effective teaching, it does attest to skills that most other faculty have not formally acquired. Dr. Tichenor has been awarded research grants and has published her pedagogical views and experiences in college science teaching journals.[1]

As Dr. Tichenor points out, "several national reports have called for reform in college science teaching (Michael 1989; AAAS Report on the National Science Foundation Disciplinary Workshops on Undergraduate Education 1990; Moore 1993; Sigma Xi 1990). The AAAS Report (1990), suggests that conventional science courses do not reflect the practice of science "at its best." The report recommends that pedagogical techniques be directed toward open-ended and investigatory laboratories in order that the teaching of science be driven by real problems rather than contrived textbook exercises and bring the spirit of scientific inquiry to undergraduate studies. The idea of student-designed laboratories supports the pedagogy suggested to foster student inquiry. I implemented these types of laboratories into my physiology course (Tichenor, 1996)."

At her university, it is required that teaching be assessed as "outstanding" for tenure to be granted. A faculty tenure committee rates teaching efficacy from (a) peer reviews, (b) student teaching evaluations, (c) yearly departmental chair reports, and (d) other evidence submitted by the candidate. Dr. Tichenor notes that, "Although the committee's final report in my case was positive overall, tenure was denied because my teaching was considered not 'outstanding.' Its decision was supported by the dean of my college." The letter from the faculty Reappointment, Promotion and Tenure committee reads:

> Dr. Tichenor has an unusual doctorate (my emphasis) focused on college teaching in the sciences and is very interested in non-traditional teaching methods. In spite of, or maybe due to this, her teaching evaluations are mixed. The review process is exacerbated when consistently mixed student evaluations are at odds with supportive statements made by peer reviewers. Often, the latter reflect peer approval of the philosophy and pedagogical approach of the instructor, while student comments address the instructor's relative success in facilitating the students' learning. Such dichotomy of opinion is especially problematic when attempting to evaluate a candidate for tenure. Contradictory statements are quite evident in this candidate's teaching evaluations. Her dean described her teaching effort as energetic, innovative and valuable to the University, but concluded that she had not distinguished herself in this preparation.

In reaching their decision, the review committee reportedly did not do a statistical analysis and comparison of her SEF with other faculty and selectively picked out certain negative comments by some students.

From the four criteria of (a) peer reviews, (b) student teaching evaluations, (c) yearly departmental chair reports, and (d) other evidence submitted by the candidate, only SEF was apparently used for the denial decision. Nearly the entire department faculty supported her and protested as a collective body the denial of tenure decision to the Dean and to the President, to no avail. A subsequent grievance committee failed to reverse the negative reappointment decision.

In commenting on her SEF responses, Dr. Tichenor points out that here evaluations in general physiology over a four-year period reflect evidence of difficulties that students have with innovative teaching:

> **Q: What do you suggest to improve this course?**
> A: 'Rely more on traditional teaching concepts than on "progressive" new teaching ideas . . '
> A. 'In my opinion, we could have learned a lot more had we stuck to a conventional lecture format.'
> A. 'Shifting to a more traditional style would be an improvement, the way it is set up now is a nice idea, but doesn't work as well as its supposed to.'
> A. 'More and better notes would be good.' and finally,
> A. 'I would suggest that there be more lecturing and less time sitting in a circle and wasting time staring at our neighbors and then when questions are asked we are told to look it up for the exam.'

On the other hand, some students who have grasped the active learning style more readily may say:

> **Q: What part of the course was most beneficial to you?**
> A. 'The fact that we had to learn through our own research, interests, etc. were most beneficial to me. I learned a lot more on my own through independent presentations than I would have merely listening to a lecture on the subjects.'
> A. 'Learning groups require people to be able to communicate with one another and therefore teach communication skills not usually taught in school. Learning groups also help people to learn problem solving skills and ways of approaching problems. Learning these skills are very important. When students leave college they are usually filled with a lot of book knowledge but not much knowledge of how to approach solving a problem. Learning groups can do this because the student is responsible for assimilating information and helping other students understand it. Also, group learning is important because it helps students learn to work in a team and take responsibility for themselves and others.'
> A. 'The part of the course most beneficial to me was the group discussions in class about the material. We are all able to relate to terms and events that enables the material to be better understood.'

Dr. Tichenor, perhaps typically, observes: "The association of mixed reviews and innovative teaching is probably familiar to many teachers who take the risk of implementing these methods. By innovative, I mean something different from the "teacher-as-textbook" model of teaching." She also, perhaps typically, that "I have found that numerical ratings for me are more positive than the open ended questions; and as it turns out, my [numerical] teaching evaluations are not at all 'mixed' but above average."

In terms of preparing students for a non lecture teaching format she has engaged in-class discussions "on primary literature, oral presentations, cases studies presented both orally and written formally, student-designed laboratories, peer evaluation, and other active learning strategies. Experiences students have disliked the most have been class periods built upon a discussion of previously assigned readings. Even though I give them a list of discussion questions prior to the class, they sometimes fail to read the discussion material."

Dr. Tichenor concludes: "there are some difficulties with the use of active learning models. Since most students have grown up with the traditional lecture method, many assume that lecturing is the only way to conduct a class. Students may feel cheated by a new approach especially if they are asked to generate their own material for class. This innovative style places new and unfamiliar demands on students. Students may feel a lack of self confidence and feel overwhelmed with the type of work expected of them."

Commentary

This case demonstrates most of the findings outlined above in this paper: (1) assumed validity of SEF, (2) reliance on SEF for administrative assessment of teaching effectiveness, (3) the subjective interpretation of SEF data by (4) untrained evaluators, both administrative and faculty, (5) reliance on SEF over peer evaluation, (6) ignoring the many variables in the implicit comparison of SEF data involved in the final decision that Dr. Tichenor was not outstanding, including (7) student bias variables, (8) selective use of qualitative written student comments, (9) over superior numerical averages on the SEF instrument, (10) justifying the interpretation of unacceptable teaching effectiveness by selective and subjective emphasis on the negative student comments in a mixed series of comments that include positive ones, (11) assumes the metaphor of student as consumer, (12) that students should have "vote" in what is appropriate teaching methods, which in term assumes, (13) students are qualified to do so, and (14) assumes that SEF validly measures teaching effectiveness.

There is an interesting set of interrelated ironies involved in this case. The university, which prides itself on being a teaching institution in apparent contradistinction to research oriented universities, denied Dr. Tichenor tenure and reappointment, a faculty who not only (a) has a rare doctoral degree that specifically prepared her for college science teaching, but (b) was engaged in what the literature suggests is one of the most effective teaching methods---collaborative and student-centered teaching----,[2] (c) was creatively trying to improve on those methods, (d) had been awarded a related grant, and (e) had published articles on her teaching in professional journals. A final irony is that Dr. Tichenor has accepted a new appointment at large state university where part of her duties will be to assess teaching.

Given the teaching orientation of the university, it would seem reasonable to expect support and encouragement for such teaching activity, especially in the light her numerical SEF score being above average. The question remains as to why this irony exists. While the answer is certainly complex, one of the answers is to retain student tuition dollars. I will deal in more detail with this issue in my final paper.

Appendix Endnotes

1. Tichenor, Linda L. (1996). Student-designed physiology laboratories: Creative instructional alternatives at a resource-poor New England university. *Journal of College Science Teaching*, *26*(3), 175-181; Tichenor, Linda L. and Joseph Kakareka. (1995). An interdisciplinary teaching approach by integrating cell biology and biochemistry: A scientific learning community at the University of New England. *Journal of College Science Teaching*, *25*(2), 144-149.

[BACK to document]

2. I might note that I am not a supporter of collaborative, student-centered teaching methods for certain populations of students such as are currently found on many colleges.

[BACK to document]

About the Author

Robert E. Haskell

Robert E. Haskell has been teaching college and university level courses for over twenty years. He earned his Ph.D. from the Pennsylvania State University in Psychology and Social Relations, his M.A., and B.A. from San Francisco State University. His areas of research and teaching include: transfer of learning, analogical reasoning, small group dynamics. Major publications include: four books, the latest of which is, *The Future of Education and Transfer of Learning: A Cognitive Theory of Learning and Instruction For The 21st Century* (forthcoming), and numerous presentations, chapters, and research articles in national and international journals. He also serves on several editorial review boards, and is Associate Editor of The Journal of Mind and Behavior. He is former Chair, and currently Professor of Psychology, Department of Social and Behavioral Sciences, University of New England.

Professor of Psychology
University of New England
Biddeford, Maine 04005

UNE Home Page: http://www.une.edu/sbs/haskell.html

E-mail: rhaskell@javanet.com

Copyright 1997 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://olam.ed.asu.edu/epaa

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, Glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692)

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Marshall University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
Arizona State University

Dennis Sayers
University of California at Davis

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina--Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher Education

Mary P. McKeown
Arizona Board of Regents

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois--UC

Robert T. Stout
Arizona State University

archives | abstracts | editors | board | submit | comment | subscribe | search

EPAA
• enter the archives • browse the abstracts • the editors • the edit board
• submit article • submit commentary • search • subscribe
volume: • 1 • 2 • 3 • 4 • 5 • 6 • 7

This article has been retrieved 2016 times since September 23, 1997.

Education Policy Analysis Archives

Volume 5 Number 19 September 23, 1997 ISSN 1068-2341

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass Glass@ASU.EDU.

College of Education

Arizona State University,Tempe AZ 85287-2411

Copyright 1997, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

Comprehensive Study of Factors Impacting Perceived Quality in School Organizations:
Findings from research on quality assessment in Iowa school districts

William K. Poston Jr.
Associate Professor
Department of Professional Studies
Iowa State University

Abstract

This paper presents the findings of studies conducted at Iowa State University of public schools in Iowa in the area of perceived quality assessment. Demographic characteristics of the respondents on the Perceived Quality Assessment Instrument from forty-four school districts were described by position, home annual income, gender, age, level of education, and years experience in current or a similar job.

The research project undertaken incorporated several studies of quality improvement characteristics of public schools. The project resulted in a compendium of coordinated research aimed at learning more about the relationships and effects of quality improvement efforts with other factors of school district operations.

The individual study components (doctoral dissertations) focused on the following issues:

- Assessment of quality improvement climate in community colleges (Bax, 1994).
- Teachers' perceptions of training programs and their relationships to total district perceptions of quality management (Johnson, 1995).
- Performance-based pay of chief executive officers and effects upon quality improvement processes in school organizations (Behounek, 1996).
- Financial characteristics of school organizations and relationships to quality management factors (Kirchoff, 1996).

This series of studies conducted over a period beginning in 1993 and culminating in 1996, was designed to assess the perceptions of school district stakeholders about the quality of their school district in terms of the seven quality dimensions of the Baldrige Award criteria. The instrument, in two versions, focused on the following Baldrige Award areas:

1. Leadership
2. Information and Analysis
3. Strategic Quality Planning
4. Human Resource Development and Management
5. Management of Process Quality
6. Quality and Operational Results
7. Client Focus and Satisfaction

Two other instruments were developed for use as a part of the -- the Staff Development and the Executive Compensation Questionnaire was developed and used in the study. Financial information was obtained from government records.

A determination was also sought to establish if the PQAI differentiated in terms of quality between high ranking and low ranking school districts. Inferential statistics established not only a significant difference between the high and low groups' quality effectiveness index, but there were also significant differences between the groups in each of the seven dimensions or sub-areas of the PQAI instrument.

A significant positive relationship was found between the perceived quality of district staff development and the perceived quality effectiveness index of the districts. Also, differences between performance-related and situational-related (nonperformance) factors were evidenced between board presidents and superintendents and a weak inverse relationship was found between performance-based compensation support and the perceived quality of the systems. Significant differences were found between the high and low QEI groups in two areas -- revenues per pupil in the leadership sub-area, and transportation cost per pupil in the information and analysis category. No correlation was found between the sample schools financial characteristics and their any PQAI rating area with one exception -- transportation cost per pupil and information and analysis.

Introduction

> *Fragmentation is the process by which we take things apart and study the parts as a means to understanding the whole. This is an appropriate model for mechanical systems, but for organizations, it is the interaction among the parts that makes things happen* (Graff, 1995).

The quality improvement emphasis of the private sector has been permeating gradually into the public schools for over a decade. The direction school organizations have been heading has been modified during this period in attempts to improve the quality of schooling, the value of educational programs, and the accomplishments of learners. Many changes were effected in schools in America during the aforementioned period, and support for sustained efforts to reform education has come from federal, state, and local levels of government (Lezotte, 1994). Once focused exclusively on manufacturing industries, quality improvement strategies have been developed and implemented in many service organizations, including school systems.

Quality improvement has become nearly a standard fixture of organizational development and staff training in schools, and the importance given to various principles of quality improvement has spawned considerable interest in the suitability of such strategies for school systems. Suddenly, policy officials are asking whether quality improvement strategies make sense for school organizations, in which terms such as "product" or "customer" have imprecise definition.

School organizations' focus on quality improvement strategies is an outgrowth of several policy efforts in the United States and other educationally progressive countries. These policy efforts indicate that everyone must become "champions of change" (Rowe, 1994). Old leadership ideas won't work anymore, and new leadership principles must recognize a number of factors:

- Organizational hierarchies don't work
- Vision has power
- Shared values create alignment
- Successful management incorporates coaching, vision, and facilitation
- Feedback on performance is crucial for improvement over time

In order to foster organizational change, educational institutions have gathered in ideas from systems theory, psychology, management theory, human-resource and organizational development, statistical process control, and human synergy. All of these ideas, in many guises and combinations, aim to remake organizations so they become more focused, disciplined, quick-footed, humane and competitive (Marchese, 1993). Many of these efforts are aimed at reducing the level of criticism toward schools and traditional types of organizational operations (Bradley, 1993)

Regardless of motivation, school systems have undertaken a number of initiatives to improve classroom instruction, academic units, staff attitudes, and financial costs (Teeter and Lozier, 1993). A plethora of school improvements and reforms have been designed and implemented with little examination of the net results or effect. Often, school reforms and improvements are focused on a subordinate part of the system rather than on system-wide change (U.S. General Accounting Office, 1993). Specific effects of quality improvement efforts have not been adequately anchored in research findings to determine their suitability and usefulness, and the need is widely apparent for identifying how and under what circumstances quality is manifested in school systems

This medley of studies examines relationships and differences among groups, factors, and conditions associated with quality improvement in schools. Specifically, this paper summarizes the problems, procedures, findings, and conclusions of four doctoral studies conducted under the supervision of Dr. William K. Poston Jr., Associate Professor of Professional Studies, Iowa State University. The four studies addressed relationships and inferences in the areas of job satisfaction and work climate, staff development and training, compensation formats and incentives, and financial characteristics. The doctoral students involved in these four study areas included Dr. Rashid M. Bax, Malaysia Ministry of Education, Dr. Pamela D. Johnson, West Des Moines (Iowa) Community School District, Dr. Thomas E. Behounek, Milo (Minnesota) Public School District, and Dr. Joseph E. Kirchoff, Manchester (Iowa) Community School District. A number of additional studies are currently underway in the areas of strategic planning, facilities and productivity, and student achievement by several other graduate students at Iowa State University.

Definitions of Terms

Quality improvement is characterized by specific terminology. Terms utilized in this study are defined below to provide clarity and understanding of their use in these studies:

> *Executive Compensation:* The sum paid to the chief executive officer of an educational institution and the underlying rationale for its determination, including results of performance or situational circumstances.
>
> *Job Satisfaction:* One's perceived emotional state expressed as a point on a scale representing the degree of positivity or pleasure resulting from the appraisal of one's job or one's experience (Locke, 1976).
>
> *Malcolm Baldrige Award:* A national award presented by the United States Government to businesses, companies, or organizations for their demonstration of excellence and exemplification of quality principles (Schenkat, 1993). It is similar to the Deming Award in Japan at least in intent.
>
> *Organizational Commitment:* Relative strength of an individual's identification with, and involvement in, a particular organization. Conceptually, it can be characterized by at least three factors: (1) a strong belief in, and acceptance of, the organization's goals and values; (2) a willingness to exert considerable effort on behalf of the organization, and (3) a strong desire to maintain membership in the organization (Porter, Steers, Mowday, and Boulian, 1974, p. 604).
>
> *Perceived Quality Assessment Instrument:* A questionnaire developed collaboratively by William K. Poston Jr. and Rashid M. Bax at Iowa State University designed to elicit perceptions of organizational stakeholders on seven dimensions of quality improvement modeled after the Baldrige Award criteria.
>
> *Perceived Quality Effectiveness Index:* A ratio or score produced by dividing the stakeholders' perceptions of the current status response score by the desired status response score, expressed as a decimal. Ratios less than one (1) indicate need for improvement; ratios greater than one (1) indicate exceeding expectations.
>
> *Quality Improvement* (synonyms: quality management, continuous quality improvement, or total quality management): a customer-focused strategic and systematic approach to continuous performance improvement (in an organization) (Vincoli, 1991, p. 28).
>
> *Staff Development or Professional Development:* Activities of school organizations that seek to prepare employees for improved performance in their present or future duties and responsibilities.
>
> *Staff Development Questionnaire:* An instrument developed for the purpose of measuring the perceived quality of staff development activities in an educational organization (Johnson, 1995).
>
> *Systems Theory:* A practice or intention to view and work with organizations as an interdependent set of components including purpose, people, methods, environment,

materials and other factors that influence the organization's functioning and activities and that affect and interrelate with all other factors.

Purposes and Rationale for the Study

There were three main purposes for conducting this series of studies. The first purpose of these studies was to determine if the Perceived Quality Assessment Instrument (commonly referred to as the PQAI) designed and developed at Iowa State University was effective in measuring factors commonly associated with excellence, efficacy, and effectiveness in school organizations. The second purpose was to identify the extent of any relationships and the nature of any associations among variables measured by the PQAI and other measures used as a part of the studies.

Finally, the purpose was to find whether or not operations and management actions in schools have an impact or influence upon the perceived level of quality within the system. Understandings obtained from these data should help school systems seeking improvement of quality to see more clearly what gaps exist among present and desired organizational conditions and what part the studied variables play in a systems approach for school organizations. The studies also attempted to demonstrate how management decisions might close the gap between the "current" and "ideal" status of factors as measured on the PQAI.

Research Questions for the Study

Questions relevant to the studies were formulated and postulated in four areas -- general perceptions of quality management, staff development, executive compensation, and financial factors. The questions addressed are listed below

1. Perceptions of Quality Management Assessment:
    - What are the perceived current and ideal levels of quality management and the quality effectiveness index (ratio) between the two levels in the selected districts?
    - What components of the perceived quality assessment instrument have the greatest impact on the overall quality rating of the selected districts?
    - What value can be ascribed to the Perceived Quality Assessment Instrument in reliability and validity in measuring system quality?
2. Staff Development Questions:
    - What are perceptions of quality of staff development programs in the selected districts?
    - Is there a relationship between the perceived quality of district staff development and the perceived level of quality management in the selected districts?
    - Is there a relationship between ratings of perceived levels of staff development programs and the seven dimensions of district quality?
    - How do districts rated highest and lowest in perceived quality of staff development differ or compare on the perceived quality effectiveness index?
3. Executive Compensation Questions
    - What factors are used and which are preferred in the compensation of the chief

executive officer of school districts by superintendents and governing board presidents and how do these two groups differ in choice of factors determining compensation?
  - How do methods used and preferred in determining compensation for the chief executive relate to levels of quality in the system as measured by the Perceived Quality Effectiveness Index?

4. Financial Factors Questions
   - To what extent are there differences between districts with perceived high and low quality on the Perceived Quality Assessment Instrument in terms of financial characteristics?
   - Is there a relationship between the financial stability or soundness of school districts and their perceived quality effectiveness index or sub-scales of the Perceived Quality Assessment Instrument?

Hypotheses and Assumptions of the Studies

All hypotheses were constructed and analyzed in null form for statistical analyses. Several assumptions supported the studies, including the following:

- The population of the study was representative of the districts in the state of Iowa, and respondents were representative of the total population of the systems involved in the studies.
- Respondents understood the content and direction of instrumentation.
- Subjects voluntarily participated in the studies by completing instrumentation.
- The measured perceptions accurately reflected actual levels of organizational quality.
- Respondents were knowledgeable about their school system, and responded accurately and honestly to all instrumentation.

Rationale for Quality Improvement in School Organizations

Continuous quality improvement, also inaccurately called total quality management, has been gathering momentum in the United States and elsewhere. The systems thinking approach was led by W. Edwards Deming, a Sioux City, Iowa native, before his death in 1993, and it has continued strongly since. The "new philosophy" has affected businesses, industries, government, and educational institutions (Brown, 1992). Literature on quality improvement has mushroomed in the past several years with application for school organizations. Deming often wrote and spoke of "continuous improvement" with the goal of quality, so the term "continuous quality improvement" is frequently used in discussions about school organization improvement, reform, or transformation (Johnson, 1995).

Education is in need of dramatic change, according to recent critics. Transformation has been called for in government, industry, and education in the United States, and such change must "be a change of state, metamorphosis, not mere patchwork on the present system of management (Deming, 1990).

Specific approaches for improvement are less than clear and distinct. According to Minnesota's Lieutenant Governor, Joanell Drystad, fragmented, individual attempts at system repair have not succeeded, and the only answer is fundamental, systematic transformation. In an address to over 2000 teachers, administrators, students, policy-makers, and business partners from 26 of the United States, Finland, and Canada, the call was made to "accelerate local transformation efforts in order to meet the national goals by the year 2000 by deploying quality quickly. By sharing and working together through the total quality systems approach, we can improve our nation's schools, provide better learning options to our students, and ensure a world-class work force" (Drystad, 1994).

A study group of forty-four school districts was identified as the population for the research project. The Perceived Quality Assessment Instrument was administered to the group of school districts. The criteria used in development of the specific items on the instrument (see Appendix for sample of the instrument) were drawn from research literature and documented successful institutional practice. The criteria reflect a number of factors that focus upon the development of educational excellence.

Continuous improvement is an organizational behavior grounded in several important characteristics -- clear goals, mission, and organizational expectations, defined direction for the design and delivery of superior teaching and learning, continuous focus on results and the use of feedback in decision-making, equitable and consistent connections among all organizational components for all clients and stakeholders, and efficient and effective use of resources (Frase, English, and Poston, 1995). These factors form a foundation for the assessment of organizational functioning, and comprise a comprehensive assessment strategy. The characteristics of such a strategy should include the following (US Dept. Commerce, 1995):

- Clear ties between what is assessed and the school system's objectives, particularly in what clients are to obtain (learning, services, etc.) or gain from the organization.
- A focus on improvement -- built upon a definition of student performance, faculty and staff capabilities, and program performance.
- Assessment as "embedded and ongoing" that is curriculum-based, criterion-referenced, and aimed at fostering improved understanding and accomplishment of goals and requirements.
- Clear guidelines as to how assessment results will be used.
- An ongoing method to evaluate the evaluation process to improve the connection between goals and client success.

For effective school improvement to occur, significant changes will have to be made in how school systems function. Many facets of school operations will have to be modified, but serious commitment calls for an understanding and implementation of quality management. The past success of quality management in business and industry raise the possibility of applying its principles to education (Teigland, 1994). It is important that the highest levels of leadership provide direction for quality improvement to occur (Bax, 1994; Walton, 1986). School administrators must seriously consider quality improvement as one option for bringing about much needed change and improvement in school organizations (Teigland, 1995).

The Quality Improvement Studies

Despite its recency in use as an organizational development tool, quality improvement in schools does not have a long history of research in its efficacy and impact upon commonly held measures of school organizational characteristics. At Iowa State University, instrumentation was developed for use in measurement of the status of perceived quality within a school organization. The assessment tool, the Perceived Quality Assessment Instrument (PQAI) was structured to determine the perceived level of quality of a school system based upon selected Baldrige Award criteria.

Baldrige Award Criteria

This group of impact studies involved an adaptation of the Malcolm Baldrige Award criteria. The Baldrige Award, begun in the United States, was designed to recognize corporations for excellence in achieving quality. The Malcolm Baldrige National Quality Award (Baldrige Award) was established in 1987 through legislation (P.L. 100-107). The purposes of the Baldrige Award were threefold:

- To promote awareness of the importance of quality improvement to the national economy;
- To recognize organizations which have made substantial improvements in products, services, and overall competitive performance; and
- To foster sharing of best practices information among U.S. organizations.

Eligibility for the Baldrige Award was initially open only to for-profit organizations. However, recent developments indicate that eligibility might be extended in the future to not-for-profit organizations.

The Baldrige Award Program strategy consists of two parts: (1) conceptual and (2) institutional. The conceptual part of the strategy involves the creation of consensus criteria which project clear values, set high standards, focus on key requirements for organizational excellence, and create means for assessing progress relative to these requirements. The institutional part of the strategy involves use of the Criteria as a basis for consistent communications within and among organizations of all types. Such communications stimulate broad involvement and cooperation, and afford a meaningful and consistent basis for sharing information. An important part of the communications is the sharing of information by Baldrige Award recipients (U.S. Dept. of Commerce, 1995).

Through the Baldrige Award, rigorous criteria were created to evaluate applicants for the Award. The Baldrige Award Criteria, based upon a set of core values and concepts, focus on key requirements for organizational excellence. These requirements are incorporated in a seven-part Criteria framework. Accompanying this framework is a set of Scoring Guidelines which permit evaluation of performance relative to the detailed Criteria. The evaluation leads to a feedback report--a summary of strengths and areas for improvement. All Baldrige Award applicants receive a feedback report.

The Baldrige Award Criteria and Scoring Guidelines have led to a number of key developments:

- Creation of a means for self-assessment;
- Replication of an award system by hundreds of organizations, including states, cities, companies, and not-for-profit organizations; and
- Creation of training programs.

Throughout the life of the Baldrige Award Program, the principal uses of the Criteria have been for such other purposes. To date, more than one million copies of the Criteria have been disseminated, and a like number of copies have been duplicated by others. This compares with a total of 546 applicants for the Baldrige Award.

Since the Baldrige Award was established in 1987, there have been 22 Award recipients (1988-1994). Award recipients have demonstrated a wide range of improvements and achievements, including product and service quality, productivity growth, customer satisfaction, reduced operating costs, and improved responsiveness. Also, Award recipients are among the Nation's leaders in investment in developing the skills of the work force.

Working Toward an Education Category Since the inception of the Baldrige Award in 1987, some educators have been involved in the Program through their service on the Award's Board of Examiners. In addition, the Award recipients have sought to involve educators and educational organizations, locally and nationally. Also, some state and local award programs already include education categories. In parallel with these Award developments, many educators have launched quality improvement efforts. National initiatives such as Goals 2000 reflect a growing national consensus to strengthen education. As a result of these and related developments, interest has grown in establishing a Baldrige Award category for education. In 1993, a decision was reached to launch Pilot activities in 1994 and 1995 to address the many issues that arise in extending eligibility to education.

Perceived Quality Assessment Instrumentation

Parallel with the development of the Baldrige Award criteria, instrumentation was under development to apply principles of the system to education. In 1993, an attempt was undertaken to develop instrumentation that focused on the factors of quality improvement in connection with school system organizational performance. The Perceived Quality Assessment Instrument was developed for use with school districts by William K. Poston Jr. and Rashid M. Bax at Iowa State University in 1993-94.

The instrument was built along the lines of the Baldrige criteria framework, and included seven dimensions: leadership, information and analysis, strategic quality planning, human resource development and management, management of process quality, quality and operational results, and client focus and satisfaction. Six to eight items were developed and included in each category. Statements were developed addressing the operations and functions of school districts, and respondents were requested to report their perception of the current situation of their system on a five point Likert-type scale. The respondents were also requested to indicate their perception of the ideal or preferred situation on the same scale on each statement. Responses included "strongly agree, agree, neutral, disagree, and strongly disagree."

Calculating the Quality Effectiveness Index

Unique to this instrument was a quality effectiveness index, which was a calculated ratio of the current situation and the ideal situation scores on each item, each of the seven dimensions, and for the total instrument. The formula for the index was expressed as follows:

$$QEI = R_c / R_i$$

QEI = Quality Effectiveness Index
Rc = Rating of Current Situation
Ri = Rating of Ideal or Preferred Situation

Calculating the quality effectiveness index in this manner permitted a basis for comparison across school systems as to the level of quality considering the ideal or favored status of the system. The expressed ratio expressed a bi-directional assessment of quality, and indicated a diminution of quality as the ratio decreased from 1.0, and an achievement of expectations as the ratio increased above 1.0. In other words, if a school system had a quality effectiveness ratio of .50, it was only approximately 50% of the way toward the level of quality it perceived as desirable. A quality effectiveness ratio of 1.25 indicates surpassing those same expectations by approximately 25% .

Such an instrument in educational research can be very valuable if carefully planned and developed (Borg and Gail, 1989). Self-assessment tools like the PQAI have been described in the professional literature as tools for obtaining an aerial view of the territory to be explored (Neuroth, Plastik, and Cleveland, 1992). By plotting a district's current location on the map, educational leaders can determine what courses of action would be appropriate for improvement.

Validation and reliability determinations for the instrument are discussed in the findings section of this report. The Perceived Quality Assessment Instrument criteria are most explicit in the areas of organizational functioning pertaining to teaching and learning. The focus on teaching and learning depends upon leadership and organizational expectations, information and analysis of assessment data for improvement, strategic planning for quality improvement and mission accomplishment, development and management of human resources congruent with participatory management, employment of process quality principles and procedures, monitoring of quality and operational results, and a clear focus on clientele and their satisfaction.

The point values and Likert-type scale represent an initial attempt to provide a basis for scoring participating school systems in terms of progress in performance improvement.

Methodology of the Studies

Population of the Study

The study involved school districts in the State of Iowa, a mid-western state with an agricultural-industrial economic climate and a rural-urban mix of communities. The specific population of the study was identified prior to the commencement of the study, so that questions could be postulated toward the respondents selected (Borg & Gall, 1989). A graduate seminar group of Ph.D. students and the major professor, common to all the students, was formed for the purpose of framing and jointly planning the study. The seminar determined several aspects of the study, including the population selected, the variables measured and scrutinized, and the activities pursuant to the completion of the research on perceived quality improvement.

Forty-four of Iowa's 360 districts were identified and selected to participate in the study by the graduate seminar group. Participation was voluntary on the part of each school organization, and efforts were made to include systems representative of different sizes and configurations. The districts selected were representative of the state as a whole in size and geographic distribution. Agreements to cooperate in the study were received from each district superintendent in order to increase the response rate. Instruments were delivered to the superintendents' offices for distribution to respondents. The superintendents were also asked to have the system's official legal representative, commonly referred to as board secretaries, select the respondents randomly and distribute the surveys.

A sample of personnel completed the study instrumentation. The Perceived Quality Assessment Instrument (PQAI) was completed by the superintendent, all board members, two administrators, five teachers, three support personnel, and two high school students for a total of 720 possible respondents in the 44 districts. The other instruments were distributed in similar manner, with different personnel responding.

Staff development information was elicited by a questionnaire designed specifically for that purpose, as was the information on compensation structures. Both the Staff Development Questionnaire and the Executive Compensation Instrument were designed specifically for this study, and both were validated with standard review and developmental procedures. Financial information used in the study was obtained from official published sources, provided by the Iowa State Department of Education.

Research Design and Variables of the Study

A survey design was used to answer the research questions. The first instrument (PQAI) consisted of two parts: Part I - Demographic Information, and Part II Rating of School System Quality Components. The other instruments were drawn from research literature. The dependent variables were the ratios (Quality Effectiveness Index) between current and ideal perceptions of quality management in each district. The specific calculation involved a ratio of the current perceived quality improvement status divided by the perceived ideal quality improvement rating. The ratios also included the seven sub-scales: Leadership, Information and Analysis, Strategic Quality Planning, Human Resource Development and Management, Management of Process Quality, Quality and Operational Results, Client Focus and Satisfaction. The independent variable of the study was the perception of quality of the districts' staff development programs. The sub-scale areas and demographic variables of the PQAI are shown in Table 1 below.

Table 1. Sub-scale areas and demographic variables of the PQAI

Quality Components	**Demographic Information**
Leadership	Position
Information and Analysis	Home Annual Income
Strategic Quality Planning	Gender
Human Resource Development and Management	Age
Management of Process Quality	Level of Education
Quality and Operational Results	Years Experience in Job
Client Focus and Satisfaction	

Development of the Instruments and Data Source

The Perceived Quality Assessment Instrument (PQAI). Use of the questionnaire in educational research can be very valuable if carefully planned and developed (Borg & Gall, 1989). In this study, three questionnaire surveys were developed for use in this study, and a fourth data-gathering technique was also used. The first survey instrument was based upon the Baldrige award criteria and labeled, "School System Perceived Quality Assessment Instrument" (Poston & Bax, 1994, see Appendix A). The instrument was based upon the seven dimensions of the Malcolm Baldrige Award areas: Leadership, Information and Analysis, Strategic Quality Planning, Human Resource Development and Management, Management of Process Quality, Quality and Operational Results, and Client Focus and Satisfaction.

Six to eight items are included in each category with statements addressing the operations and policies of school districts. Respondents were asked to judge their current situation and the desired or ideal situation in their school system for each item using a Likert scale (Borg & Gall, 1989) of five possible responses for each (Strongly Disagree, Disagree, Neutral, Agree, Strongly Agree). Demographic items were also included (position, income, gender, age, level of education, and years of experience in current or similar job) to aid in possible statistical comparisons and analyses of the groups. The instrument was refined and improved following Bax's study (Bax, 1994).

The Perceived Quality Assessment Instrument (PQAI) was developed and based upon principles of sound informational and research principles. Borg and Gall (1989) report that specific behaviors can be predicted from attitude measures about those behaviors, therefore a

Likert-type scale of five possible responses for each item was again used: Almost Never, Occasionally, Don't Know, Frequently, Almost Always (Knudsen, 1993). The third response, Don't Know, was included in the third position to encourage either a negative or positive answer from respondents. Questions posed in closed form aid in the efficient quantification and analysis of results (Borg & Gall, 1989). Some items were reversed from the others, so that not all statements would be written in similar (positive) format (Borg & Gall, 1989), and (2) so that respondents would not feel that every answer should be marked Almost Always to be "correct." The length of the items and the length of the questionnaire itself were kept as short as possible for ease in understanding and increased chance of the instruments being returned (Borg & Gall, 1989). In addition, several devices were used to encourage timely participation, including a self-addressed postage-paid mailer, phone calls, and personal contacts.

The PQAI was validated by thirteen experts in quality improvement who were educational leaders, superintendents, university professors, or university officials with an understanding of quality processes. The panel is listed in the Appendix. Each member of the panel was asked to assure that each item accurately reflects the concepts purported to be measured by the instruments, to evaluate the items for clarity and completeness, and to make suggestions for improvement. Based on the recommendations of the panels, the instruments were revised. The Perceived Quality Assessment Instrument was further pilot tested by a group of Iowa State University doctoral students studying quality management in public schools. The PQAI instrument took approximately ten to fifteen minutes to complete.

Development of the Other Instruments and Data Sources

Staff Development Instrumentation. The District Staff Developmental Questionnaire (Johnson, 1995) was developed by one of the members of the study group (Johnson, 1995) after reviewing staff development literature in the following areas: (1) planning, (2) administrative support, (3) delivery, (4) follow-up, and (5) evaluation. (Wood et al., 1982; Fullan, 1985; Joyce & Showers, 1988; Sparks and Loucks-Horsley, 1989; Senge, 1990; Rebore, 1991; Guskey, 1994; McBride et al., 1994; Sparks & Vaughn, 1994) Orlich, 1983; Guskey & Sparks, 1991; Templeman & Peters, 1992; Showers, 1985; Wade, 1985; Bates & Stachowski, 1991; Wood & Thompson, 1993). The instrument was validated by ten staff development professionals recognized throughout the state and nation as very knowledgeable in this field. A Likert scale approach was used for the instrument.

Executive Compensation Instrumentation. The Current and Preferred Methods of Awarding Superintendent Salary Increases Questionnaire (Behounek, 1996) survey instrument was developed following a review of related literature and an interview of ten superintendents and twelve board members about methods of awarding compensation to superintendents. Twelve criteria were incorporated in the instrument, with five of the criteria based upon the superintendent's performance and seven of the criteria nonperformance related. The instrument was validated by a panel of nine practicing administrators and researchers, and field tested with twelve board members. A Likert-type scale was used for the instrument.

Financial Information Instrumentation and Data Sources. The source of financial information for the selected Iowa school districts was obtained from the official financial audits for each system from the 1992-93 school year, the most recent data available (Kirchoff, 1996). The source of the audits used in the study was the State Auditor's Office, a division of government in Iowa. The audits were used as the source of student enrollment, fund balances, revenues and expenditures.

Human Subjects Committee Approval

The Iowa State University Committee on the Use of Human Subjects in Research reviewed all research projects as a part of this study and concluded that the rights and welfare of the human subjects were adequately protected, that risks were outweighed by the potential benefits and expected value of the knowledge sought, that confidentiality of data was assured, and that informed consent was obtained by appropriate procedures.

All instruments employed the use of anonymity for respondents, as there was no request for names of respondents. However, the questionnaires were coded to identify which school system the respondents represented. This was necessary for follow-up and data analysis. The completion of the questionnaires was voluntary and constituted consent to participate in the research project. All questionnaires were kept secure throughout the duration of the study and filed for safekeeping after the study was concluded.

Data Collection Procedures

Study instrumentation was personally delivered to each superintendent or his/her designee in each district in late February and early March of 1994. The accompanying directions for completing the surveys were self-explanatory. Instructions were printed on the fronts and backs of the questionnaires to mail the completed instruments directly to Iowa State University by dropping them in the U.S. mail. A postal permit was printed on the back of each. To increase the rate of return, phone calls were made to participating superintendents in April asking them to encourage those who had not yet responded to do so.

As a result, 471 of the 720 PQAI surveys were received for a return rate of 65.4%.

Statistical Analysis of The Data

Perceived Quality Assessment Instrument. After the surveys were returned, the responses were entered into a Microsoft Excel spreadsheet program. Only one of the PQAI instruments was discarded because it was damaged in the mail. Four of the districts were removed from the study because an insufficient number of PQAI surveys were returned (one survey and was returned from two districts and three surveys were returned from two others). As a result, 462 surveys representing forty districts were used for data analysis. The items were ranked on a scale ranging from 1 (strongly disagree) to 5 (strongly agree). Means for each respondent and each district for each question were computed. On the Perceived Quality Assessment Instrument, means for each respondent and district for the current and ideal situations were calculated as well as the ratio (quality effectiveness index) of the current divided by the ideal.

Descriptive statistics were computed on the demographic variables of the PQAI. Preliminary analyses using the one-way analysis of variance (ANOVA) were conducted to compare the responses of demographic groups. Factor analysis was conducted on the PQAI to determine reliability. The ANOVA and Scheffe' post-hoc method were used to determine differences between the four groups.

Other instruments. Spearman rho and Pearson product-moment correlation coefficients were computed to test for a relationship between responses on the PQAI and the staff development and compensation instruments and between financial factors and the PQAI. Stepwise multiple regression was used to find a relationship between staff development and the seven PQAI dimensions. The differences between high and low districts on the compensation and staff development instruments and financial factors were determined using the t test and analysis of variance techniques.

The Cronbach alpha reliability coefficient was also calculated for the Staff Development

Questionnaire and the Executive Compensation Instrument to determine the internal consistency of the total instruments.

Findings of the Studies

Introduction

The findings of the study are organized into the following sections: (1) General Characteristics of the Sample, (2) Reliability Analysis of the Instruments, (3) Quality Management Perceptions, (4) Staff Development Perceptions, (5) Results of Hypotheses Tested, (6) Evaluation of the Perceived Quality Assessment Instrument, and (7) Summary.

General Characteristics of the Sample

The primary purpose of this section is to describe the participants in this study who completed the Perceived Quality Assessment Instrument (PQAI) with respect to the following demographic variables: (a) position, (b) home annual income, (c) gender, (d) age, (e) level of education, and (f) years experience in current/similar job. The descriptive information is presented in Table 2 which follows.

> ***Position*** - Of the 458 respondents who filled in this question, 132 (28.8%) were teachers, 76 support staff (16.6%), 68 administrators (14.8%), 30 superintendents (6.6%), 108 board members (23.6%), and 44 students (9.6%). Four people did not complete this category.
>
> ***Home Annual Income*** - Of the 358 people answering this question, 24 reported earnings of less than \$10,000 (6.7%), 77 from \$10,000 to \$29,999 (21.4%), 147 from \$30,000 to 49,999 (40.9%), and 110 reported earnings of \$50,000 or more (30.9%). There were 104 respondents who did not fill in this information.
>
> ***Gender*** - The number of male and female participants reporting their gender was fairly even. Male respondents numbered 215 (52.6%) while females totaled 194 (47.4%). The number who did not fill in this item was 53.
>
> ***Age*** - The age of the respondents was divided into five categories: (1) under 18, (2) 18-29, (3) 30-55, (4) 56-70, and (5) over 70. The largest group was the 3055 year category (73.8%). The second largest group was the 56-70 year group (11.9%). The smallest group was the over 70 category with only one person. Forty-two people did not report their age group.
>
> ***Level of Education*** - The educational level of the respondents was classified into four categories: (1) less than a B.A. degree, (2) B.A. degree, (3) Master's degree, and (4) Doctorate degree. The largest group (151 people) represented those who had less than a B.A. degree, (34.6%). Those having a B.A. degree were the second largest group with 142 people (32.5%). The third largest group, those having a Master's degree, were very close with 129 people ((29.5%). Fifteen respondents (3.4%) reported having a Doctorate degree. Twenty-five respondents did not report their level of education.
>
> ***Years of Experience*** - The years of experience of the respondents were divided into

four categories: (1) under 5 years, (2) 5-10 years, (3) 11-25 years, and (4) 25 years or more. The findings revealed that 191 (44.3%) of the participants had 11-25 years of experience, 96 (22.3%) had 5-10 years of experience, 88 (20.4%) had under five years, and 56 (13%) had 25 or more years of experience. Thirty-one respondents did not complete this item.

Table 2. Demographic information of respondents on the PQAI

Category		Frequency	Percent
Position	Teacher	132	28.8
	Support Staff	76	16.6
	Administrator	68	14.8
	Superintendent	30	6.6
	Board Member	108	23.6
	Other	44	9.6
	Total	458	100.0
Income	< $10,000	24	6.7
	$10,000 - $29,999	77	21.4
	$30,000 - $49,999	147	40.9
	$50,000 +	110	30.9
	Total	358	100.0
Gender	Male	215	52.6
	Female	194	47.4
	Total	409	100.0
Age	< 18	23	5.5
	18 - 29	36	8.6
	30 - 55	310	73.8
	56 - 70	50	11.9
	70 +	1	0.2
	Total	420	100.0
Education	< Baccalaureate	151	34.6
	Baccalaureate	142	32.5
	Master's Degree	129	29.5
	Doctorate	15	3.4
	Total	437	100.0
Experience	< 5 years	88	20.4
	5 - 10 years	96	22.3
	11 -25 years	191	44.3
	26 + years	56	13.0
	Total	431	100.0

One-way analysis of variance and t-test procedures were used to determine differences in the quality effectiveness index among different demographic groups. It was found that there were significant differences in responses by position, gender and education. The Scheff© method was used to analyze these differences and established that there were significant differences in the position category between the responses of teachers and administrators and between teachers and board members, both at the .05 level. A description of respondents by position follows.

Analysis by position

In order to gain a better understanding of the respondents on the PQAI and their differing characteristics, further analysis was conducted on the position category by generating cross-tabulations with the other variables--income, gender, age, education and experience. The results are shown in Table 3 that follows.

Position with income - As shown in the following Table , the majority of teachers (50) reported incomes in the $30,000 to $49,999 category with the second highest number (44) being at the over $50,000 level. Support staff reported 29 people in the $30,000 to $49,999 and 27 at the $10,000 to $29,999 level.

For administrators, the two highest categories were over $50,000 (39) and $30,000 to $49,999 (28). Only one administrator reported making less than $29,999. All of the superintendents except two (27) reported incomes over $50,000, while the remaining two reported earnings in the $30,000 to $49,999 category. The majority of board members (51) were in the over $50,000 category, with 31 at the $30,000 to $49,999 level, and 16 at the $10,000 to $29,999 level. Of those in the other category (students), 23 reported earnings under $10,000 with the remainder (16) dispersed throughout the other three levels.

Table 3. Demographic information of PQAI respondents by position

Category	Teacher	Support Staff	Admin.	Supt.	Board	Other	Total	Percent
Income (35 missing)								
< $10,000	0	1	0	0	0	23	24	5.6
$10,000 - $29,999	30	27	1	0	16	3	77	18.0
$30,000 - $49,000	50	29	28	2	31	7	147	34.4
$50,000 +	44	12	39	27	51	6	179	41.9
Total	124	69	68	29	98	39	427	100.0
Gender (53 missing)								
Male	39	21	50	22	65	18	215	52.6
Female	82	48	10	0	30	24	194	47.4
Total	121	69	60	22	95	42	409	100.0
Age (42 missing)								
< 18	1	1	0	0	0	21	23	5.5
18 - 29	18	4	1	1	0	12	36	8.6
30 - 55	93	56	54	22	80	5	310	73.8
56 - 70	10	9	8	6	14	3	50	11.9
70 +	0	0	0	0	1	0	1	.2
Total	122	70	63	29	95	41	420	100.0
Education (25 missing)								
< Baccalaureate	1	57	1	0	55	37	151	34.6
Baccalaureate	95	10	3	2	32	0	142	32.5
Masters Degree	0	0	0	9	6	0	15	3.4
Doctorate	0	0	0	9	6	0	15	3.4
Total	126	71	68	29	403	40	437	100.0
Experience (31 missing)								
< 5 years	12	13	18	4	17	24	88	20.4
5 - 10 years	30	21	17	8	20	0	96	22.3
11 - 25 years	67	33	24	15	48	4	191	44.3
26 + years	17	6	9	2	18	4	56	13.0
Total	126	73	68	29	103	32	431	100.0

Position with gender - There were over twice as many females as males in the teacher (82 and 39) and support staff (48 and 21) categories, as shown on Table 3.

However, male administrators (50) outnumbered females (10) 5 to 1, while there were no female superintendents and 22 male superintendents who reported their gender. There were more than twice as many male board members (65) as females (30). In the other category, 18 were males and 24 were females.

Position with age - The majority of the teachers (93) were in the 30-55 years age group, with 18 falling in the 18-29 category, 10 in the 56-70 division, and one reporting being under 18. Most of the support staff (56) reported being 30-55, with 9 in the 56-70 age group, 4 in the 30-55 range, and one under 18. Administrators, too, reported the majority (54) in the 30-55 group, with 8 in the 56-70 category, and one in the 18-29 age group. Twenty-two superintendents reported being in the 30-55 age range, with six in the 56-70 category, and only one in the 18-29 age group. No board members reported being under 30, while the vast majority (80) were in the 30-55 range, 14 in the 56-70 category, and 1 over 70. In the other group, 21 reported being under 18, 12 said they were 18-29, 5 filled in the 30-55 group, and 3 marked the 5670 age group.

Position with education - Most of the teachers (95) reported having B.A. degrees, with 30 having Master's degrees, and one reporting less than a B.A. The majority of support staff (57) had less than a B.A. degree, with 10 having a B.A., and 4 having a master's degree. Most of the administrators (62) had a master's degree, with only 3 listing a B.A., and one filling in the less than B.A. category. Nearly two-thirds of the administrators (18) had master's degrees, while nine had doctorates, and two had completed their B.A.'s. Of the board members reporting, over half (55) had less than a B.A., while 32 had a B.A., 10 had a master's degree, and six held doctorates. In the other group, 37 reported having less than a B.A., while three filled in the master's degree response. It would appear that there is an inverse relationship between the education and income levels of the board members. While the majority of board members did not have a college degree, over half of them reported incomes of more than $50,000.

Position with experience - The results from Table 2 show that the largest number of respondents in every group except other had 11-25 years experience in the current or similar job. The second highest level in every group except administrators and other was the 5-10 year category. The remainder of respondents were distributed fairly evenly between the other two categories.

Analysis of other demographics

The ANOVA procedure revealed no significant differences on the quality effectiveness index in income levels, age, or experience. However, differences were found for education levels. Using Scheff©, significant differences were found between those with less than a B.A. and those with a B.A. degree. No other significant differences were found between the other education levels.

Although the t test indicated a significant difference in the quality effectiveness index between males and females, when t tests were conducted for gender by position, no significant differences were found in any position level for gender. Thus the difference was only present when all levels were considered as a whole.

Reliability Analysis of the Instruments

The SPSS program was utilized to determine the reliability of the Perceived Quality Assessment Instrument. Analyses were conducted for the current and ideal sections and for the overall scale (total instrument). The alpha reliability coefficients are reported in the Table which follows. The alpha coefficients ranged from .68 to .85 for the current section of the instrument, with an overall reliability of .96. The alpha coefficients were somewhat different for the ideal section, ranging from .61 to .89, with an overall reliability of .94, indicating a high positive correlation among all items.

Table 4 below presents the information and reliability coefficients for the PQAI

Instrument.

Table 4. Reliability analysis of current and ideal sections of the PQAI

	Reliability				
PQAI Dimension	Item Numbers	Current	N	Ideal	N
Leadership	1 - 6	.84	454	.82	451
Information and Analysis	7 - 12	.85	443	.82	430
Strategic Quality Planning	13 - 18	.84	451	.87	445
H.R. Development & Mgmt.	19 - 25	.83	451	.61	441
Mgmt. of process Quality	26 - 31	.79	440	.85	437
Quality & Operational Results	32 - 37	.68	454	.75	448
Client Focus & Satisfaction	38 - 45	.85	453	.89	444
Overall	1 - 45	.96		.94	

The Cronbach alpha reliability coefficient was also calculated for the Staff Development Questionnaire to determine the internal consistency of the total instrument. Estimates of internal consistency are based on the average correlation among items within a test or instrument. The reliability coefficient for all thirty items was .95, again a high positive correlation. This level of reliability was determined to be more than sufficient for the purposes of this research.

On the Compensation Assessment Instrument, the Cronbach alpha reliability coefficient was also calculated to determine an estimate of the internal consistency of the instrument. Reliability coefficients were calculated for the two sets of questions -- one for current practices and the other for ideal (or preferred) practices in compensating superintendents, or school system chief executive officers. Table 5 below contains the reliability coefficients for the four sub-areas and the total instrument (Behounek, 1996).

Table 5. Cronbach Alpha reliability coefficients for sub-areas of the Compensation Assessment Instrument

Variable	Alpha	Number of Items
Current Practice		
Non-performance related items	.75	7
Performance related items	.83	5
Ideal (Preferred) Practice		
Non-performance related items	.87	7
Performance related items	.67	5
Total of all items	.67	24

The level of reliability of the Compensation Assessment Instrument was determined to be adequate, albeit modest, for the purposes of this research (Nunnally and Durham, 1975).

Quality Management Perceptions

The study sought to determine the perceived current and ideal levels of quality management in each participating district and the Quality Effectiveness Index ratio between the two assessments of perception. Means were first calculated for the responses for each item for each district. The means of all of the current and ideal responses for each district were then figured. Current means ranged from 2.80 (district HH) to 3.92 (district MM). Ideal means ranged from 3.77 (LL) to 4.56 (P and R).

Finally, the ratio between the current and ideal means for each district were determined. The quality effectiveness index ranged from .6532 (district HH) to .9506 (district A). Four districts were removed from the study because of their low return (N for districts F and V = 1 survey each; N for districts I and M = 3 surveys each). It was felt that such a small sample from those districts would not give a reliable representation of the perceptions of the entire district. The presentation in Table 6 below shows the distribution of means for current and ideal situations and the ratios for the remaining 40 districts.

Table 6. Current and ideal means and quality effectiveness index for districts on the Perceived Quality Assessment Instrument

District Code	Rank	Current Situation Mean	Ideal Situation Mean	Quality Effectiveness Index
A	1	3.60	3.79	.9506
AA	23	3.10	4.00	.7747
B	10	3.52	4.18	.8408
BB	33	3.44	4.66	.7371
C	7	3.70	4.32	.8572
CC	31	3.32	4.53	.7322
D	36	3.10	4.36	.7119
DD	15	3.39	4.23	.8005
E	30	3.08	4.15	.7419
EE	32	2.97	4.01	.7399
FF	4	3.66	4.15	.8859
G	21	3.44	4.37	.7865
GG	16	3.30	4.12	.8001
H	14	3.24	4.04	.8020
HH	40	2.81	4.30	.6532
II	28	3.22	4.29	.7500
J	13	3.45	4.26	.8083
JJ	11	3.56	4.31	.8250
K	19	3.33	4.19	.7943
KK	38	2.97	4.30	.6899
L	27	3.26	4.32	.7554
LL	5	3.50	3.95	.8843
MM	6	3.92	4.46	.8784
N	26	3.43	4.50	.7619
NN	22	3.40	4.42	.7703
O	3	3.77	4.23	.8909
OO	2	3.71	4.08	.9107
P	39	2.99	4.56	.6545
PP	8	3.64	4.29	.8501
Q	35	3.20	4.36	.7357
QQ	18	3.58	4.47	.7989
R	29	3.42	4.58	.7483
RR	20	3.56	4.50	.7889
S	25	3.23	4.23	.7628
T	24	3.26	4.25	.7669
U	9	3.46	4.09	.8480
W	34	3.27	4.44	.7369
X	17	3.44	4.31	.7990
Y	37	3.38	4.28	.7097
Z	12	3.65	4.44	.8234

Staff Development Perceptions and PQAI

To determine the perceptions of teachers in the participating school districts about their systems staff development programs, means for each question for each district were calculated. Then the mean of all the responses for each district was determined. These means ranged from 1.963 to 3.789. The same four districts (F, I, M, and V) were removed from the list for comparisons with the group used with the PQAI. Ranks of districts on the Staff Development Questionnaire were compared with the districts' ranks on the PQAI, as shown in Table 7 below.

Table 7. Comparison of ranks on the Staff Development Questionnaire and the Perceived Quality Assessment Instrument

District Code	SDQ Rank	PQAI Rank	Sub-Group
A	1	1	1
K	2	19	1
PP	3	8	1
J	4	13	1
JJ	5	11	1
LL	6	5	1
QQ	7	18	1
L	8	27	1
C	9	7	1
II	10	28	1
O	11	3	2
Q	12	26	2
OO	13	2	2
DD	14	15	2
H	15	14	2
B	16	10	2
RR	17	20	2
P	18	39	2
R	19	29	2
MM	20	6	2
N	21	26	3
BB	22	26	3
Z	23	16	3
GG	24	16	3
NN	25	22	3
AA	26	23	3
W	27	34	3
U	28	9	3
D	29	36	3
T	30	24	3
X	31	17	4
S	32	25	4
Y	33	37	4
EE	34	32	4
E	35	30	4
CC	36	31	4
G	37	21	4
FF	38	4	4
HH	39	40	4
KK	40	38	4

The forty districts, listed above, were divided into four groups for further analysis. The sub-groups were comprised of groups of ten districts each, according to their rank on the Staff Development Questionnaire. A one-way analysis of variance was calculated to determine differences in means of the four sub-groups. The results of the ANOVA are shown below in Table 8 below.

Table 8. One-way analysis of variance of group means on the Staff Development Questionnaire

Source	df	SS	MS	F	F_{cv}
Between groups	3	7.53	2.51	129.53	4.51
Within groups	36	.70	.02		
Total	39	8.22			

$p= <.0001$

The four groups appeared to differ in terms of their group means on the Staff Development Questionnaire, as illustrated in the Table (Table 9) below.

Table 9. Sub-group means and standard deviations on the Staff Development Questionnaire

Group	N	Mean	St.Dev.
1	10	3.36	.2129
2	10	2.93	.0720
3	10	2.65	.0859
4	10	2.17	.1399

A one way analysis of variance was then employed to determine differences in means between and among the four sub-groups. Significant differences were found among the groups, and a Scheff© method was used to determine which sub-groups differed from one another. All groups were found to differ from each other significantly, as shown in Table 10 which follows:

Table 10. Comparison of group means on the Staff Development Questionnaire

Sub-Group Comparison	Mean Difference	Scheffe *F* test
1 vs. 2	.4366	16.4*
1 vs. 3	.7155	44.05*
1 vs. 4	1.1944	122.76*
2 vs. 3	.2789	6.69*
2 vs. 4	.7578	49.42*
3 vs. 4	.4789	19.74*

*Significant at .01 level

PQAI and SDQ Relationships

The relationship between the PQAI measurement of perceived quality and the Staff Development Questionnaire was explored. The composite district ratings of staff development programs were correlated with the Quality Effectiveness Index computations. A Spearman rho test was used to calculate correlations between the two indices. A correlation of .50131 was found, and it was significant at the .01 level of confidence. This relationship was determined to be moderate, but significant.

A second exploration was conducted to determine if there were any relationships between ratings of perceived levels of staff development quality and the seven dimensions of district quality (current) measured by the PQAI in the participating districts.

To accomplish this purpose, means of the current perceptions in each of the seven quality dimensions of the Perceived Quality Assessment Instrument for the ten highest ranking districts on the Staff Development Questionnaire were computed. The mean of each district on the Staff Development Questionnaire was compared with the mean of each quality dimension for these districts. Stepwise multiple regression was used to determine the relationship of the measured quality of staff development to the seven quality dimensions. The seven variables were entered one at a time and a significance test was conducted to determine the contribution of each (Hinkle, Wiersma, & Jurs, 1988). The stepwise solution was terminated when the remaining variables did not make a statistically significant contribution to the regression.

It was found that three of the seven dimensions--Client Focus and Satisfaction, Quality and Operational Results, and Management of Process Quality--were good predictors of levels of quality staff development. The adjusted R2 (squared) was .956 indicating that 96 percent of the variance on the Staff Development Questionnaire was explained by these three dimensions. The remaining four dimensions -- Leadership, Information and Analysis, Strategic Quality Planning, and Human Resource Development and Management -- did not predict levels of district staff development in any significant degree.

Correlation matrices were also constructed in order to show the interrelationships between all the sub-scales on the Perceived Quality Assessment Instrument. Table 11 shows the correlation coefficients for each of the seven sub-scales. As a rule of thumb, correlation coefficients between 00 and 30 show little if any correlation; 30 to 50, a low correlation; 50 to 70, a moderate correlation; 70 to 90, a high correlation; and .90 to 1.00, a very high correlation (Hinkle, Wiersma, & Jurs, 1988).

Using these guidelines, the matrix shows that the majority of the correlations are moderate positive correlations. None of the relationships are below 50, with the lowest being between Leadership and Quality and Operational Results (.52).

Five of the relationships were in the high positive category: Human Resource Development and Management and Strategic Quality Planning (.77), Human Resource Development and Management of Process Quality (.73), Management of Process Quality and Quality and Operational Results (.71), Client Focus and Satisfaction and Management of Process Quality (.72), and Client Focus and Satisfaction and Quality and Operational Results (.74). These relationships show that all seven dimensions are related to each other and measure the same concept -- district quality management.

Table 11. Correlation matrix for the seven dimensions of the PQAI

PQAI Sub-scale	**Leadership**	**Infor. & Analysis**	**Strategic Planning**	**H.R. Dev. & Mgt.**	**Process Quality**	**Quality Results**	**Client Focus**
Leadership	1.00						
Information	.63	1.00					
Planning	.61	.69	1.00				
H.R. Mgt.	.62	.67	.77	1.00			
Process	.63	.69	.67	.73	1.00		
Results	.52	.62	.62	.67	.71	1.00	
Focus	.60	.64	.64	.69	.72	.74	1.00

Still another exploration was conducted into the area of determining whether or not districts ranked highest or lowest in perceived staff development quality differed significantly when measured by the perceived quality effectiveness index.

Table 12 shows these highest and lowest districts, their ranks on both instruments and their mean scores for each.

Table 12. Mean scores and ranks for the highest and lowest ten districts on the Staff Development Questionnaire and the Perceived Quality Assessment Instrument

Table 12. Mean scores and ranks for the highest and lowest ten districts on the Staff Development Questionnaire and the Perceived Quality Assessment Instrument

District	Rank on SDQ	Group	SDQ Mean	PQAI Mean	Rank on PQAI
A	1	1	3.789	.9506	1
K	2	1	3.567	.7943	19
PP	3	1	4.442	.8501	8
J	4	1	3.427	.8083	13
JJ	5	1	3.408	.8250	11
LL	6	1	3.278	.8843	5
CD	7	1	3.278	.7989	18
L	8	1	3.273	.7554	27
C	9	1	3.113	.8572	7
II	10	1	3.067	.7500	28
X	31	4	2.408	.7990	17
S	32	4	2.310	.7628	25
Y	33	4	2/280	.7097	37
EE	34	4	2.211	.7399	32
E	35	4	2.203	.7419	30
CC	36	4	2.167	.7322	31
G	37	4	2.150	.7865	21
P	38	4	2.042	.8859	4
HH	39	4	1.994	.6532	40
KK	40	4	1.963	.6899	38

Using the unpaired t test on the means of the two groups on both instruments, significant differences were found for the highest and lowest groups on the Perceived Quality Assessment Instrument. A correlation coefficient (Pearson product-moment) was then calculated among groups compared on both instruments. A moderate positive correlation (.564) was found for the top ten districts which was significant at the .05 level. The correlation between the bottom ten districts on both instruments (.264) was not significant. However, when district FF was removed which ranked high (number 4) on the Perceived Quality Assessment Instrument, the correlation rose to .710, a high positive correlation. It was determined that the higher a system demonstrated perceived quality in staff development, the higher its perceived quality effectiveness index.

Analysis of Executive Compensation Perceptions and PQAI

The compensation of executives and leaders has often been thought to be instrumental in influencing the level of performance on the job. In short, some believe that money can motivate (Poston and Frase, 1992). Deming, long considered the "father of quality management," however, felt that the theory of psychology eroded any confidence that money could motivate individuals on the job (Deming, 1986). His notion was that money and employee compensation had little to do with the quality of an organization.

To test that theory in part, this study undertook to examine the perceptions of school board members and superintendents about the effect or influence of salary compensation upon the quality of an organization. Perceptions from board members and superintendents were elicited with the instrument earlier described. The following Table demonstrates the tabulated results of the perceptual assessment.

Table 13. Aggregate responses of superintendents and board presidents concerning practices and preferences for salary increases

		Board Presidents			Superintendents		
Item	**Belief statement**	***N***	**Mean**	**St.Dev.**	***N***	**Mean**	**St.Dev.**
CURRENT: Unrelated to Performance							
1	Length of service to system	44	2.67	.94	44	2.68	.93
2	Increased formal education	44	2.75	.97	44	2.27	.95
3	Cost of living adjustment	44	2.86	1.00	44	2.48	1.09
4	Growth in revenues	44	3.46	1.15	44	3.50	1.07
5	Settlement of teacher contract	43	3.56	1.22	44	3.80	1.17
6	Regional colleagues status	44	3.39	.99	44	3.48	1.07
7	Student enrollment change	44	1.86	1.02	44	2.16	1.16
	Total	44	2.94	1.17	44	2.91	1.25
CURRENT: Related to Performance							
8	Performance on job	44	3.96	1.16	44	3.73	.85
9	Achievement of set goals	44	3.61	1.13	44	3.11	.99
10	School Board evaluation	44	3.77	1.05	44	3.61	.99
11	Student achievement	44	2.77	1.14	44	2.02	1.02
12	Test score results	44	2.05	1.12	44	1.73	.90
	Total	44	3.23	1.32	44	2.84	1.25
IDEAL: Unrelated to Performance							
1	Length of service to system	44	2.55	1.02	44	2.97	1.05
2	Increased formal education	44	3.41	.79	44	3.66	.78
3	Cost of living adjustment	44	3.41	.79	44	3.41	1.00
4	Growth in revenues	44	3.02	1.00	44	3.48	.95
5	Settlement of teacher contract	44	3.46	1.07	44	3.21	1.10
6	Regional colleagues status	44	3.07	1.13	44	4.18	.62
7	Student enrollment chnage	44	2.46	.79	44	2.59	.87
	Total	44	3.07	1.04	44	3.36	1.03
IDEAL: Related to Performance							
8	Performance on job	44	4.36	.72	44	4.59	.50
9	Achievement of set goals	44	4.18	.79	44	4.32	.86
10	School Board evaluation	44	4.11	.75	44	4.43	.50
11	Student achievement	44	3.21	.85	44	2.84	1.08
12	Test score results	44	2.55	.82	44	2.02	.88
	Total	44	3.68	1.05	44	3.64	1.29

For the purposes of this study, the strength of agreement indicated by each level of response was categorized as low if less than 2.25, moderate if between 2.26 and 3.75, and high if more than 3.76. These designations were selected arbitrarily for purposes of comparison only. The results in Table 13 represent superintendent and board president perceptions on how increases for superintendents are awarded in current practice and how they should be awarded given conditions of ideal practice.

Board presidents and superintendents agreed that increases in enrollment have little to do with compensation, but levels of compensation awarded pursuant to teacher contract deliberations have a high impact upon the level of superintendent's compensation. In the area of performance-related factors, the highest factor of influence to both board presidents and superintendents was the superintendent's performance on the job. The lowest relationship to compensation was perceived to be with standardized test results. Ironically, this measure of organizational effectiveness had low agreement as to its importance in determining compensation.

When considering the ideal situation, board presidents and superintendents agreed that on-the-job performance, achievement of district goals, and board evaluation of the superintendent should have the greatest level of influence on the superintendent's compensation. Ideally, as in the current status, student test performance was perceived of low importance. In non-performance rated ideal conditions, superintendents favored most strongly the use of "benchmarking" salaries of peers or colleagues in the region for use in determining the salary of the superintendent. Changes in enrollment fared no better in perceptions of ideal factors than with current conditions, as it was rated as the lowest in agreement for both groups.

To more adequately compare board presidents' and superintendents' perceptions about compensation, the mean scores of both groups wére subjected to analysis. Unpaired, two-tailed t-tests were used to determine if significant differences existed between the two groups on the four areas of the study. Responses from superintendents and board presidents were also compared between performance and non-performance factors and current practice and ideal situation. In these cases, paired, two-tailed t-tests were utilized. Significance was set at the .05 level of confidence for the purpose of this analysis.

Table 14. t-Test results of group means for determining superintendent compensation

Group	N	St. Dev.	Mean	*t*	P
Current: Unrelated to Performance					
Board Presidents	44	.50	2.93	.18	.857
Superintendents	44	.71	2.91		
Current: Related to Performance					
Board Presidents	44	.90	3.23	2.2	.030
Superintendents	44	.76	2.84		
Ideal: Unrelated to Performance					
Board Presidents	44	.49	3.07	-2.70	.008
Superintendents	44	.51	3.36		
Ideal: Related to Performance					
Board Presidents	44	.46	3.68	.40	.690
Superintendents	44	.50	3.64		

In the area of perceptions of board presidents and superintendents considering the importance of non-performance related job factors, the analysis revealed that there was no significant difference between the two groups. On the other hand, when looking at how they two group means compare on performance-related job factors, there was a significant difference between board presidents and superintendents in their perceptions. Board presidents felt more strongly about using performance in the determination of compensation for the chief executive.

Two more tests of significant differences were employed, dealing with perceptions in the ideal or preferred situation. When looking at the importance of non-performance ideal factors, board presidents placed less credence on these factors than did superintendents. The difference was significant, indicating that superintendents placed more importance on non-performance job factors in determining superintendent compensation than did board members in an ideal situation.

In the ideal situation using performance-related compensation factors, superintendents and board members were not significantly in disagreement. Both groups rated the use of performance factors in determining the level of job compensation in the "high-moderate" range.

To compare performance with non-performance factors across groups, a series of analyses employing t-tests were conducted. Superintendents' and board presidents' perceptions on performance against nonperformance factors in current practice and performance against nonperformance factors in ideal practice were analyzed. Comparisons of superintendent and board president perceptions of current practices against ideal practices across performance and non-performance areas were also examined. Results of these analyses are displayed in Table 15 which follows.

Table 15. T-Test results between groups on performance and non-performance factors in determining superintendent compensation

Group	N	St. Dev.	Mean	*t*	P
Current: Superintendents					
Non-performance factors	44	.71	2.91	.58	.5636
Performance factors	44	.76	2.84		
Current: Superintendents					
Non-performance factors	44	.50	3.36	-2.46	.0180
Performance factors	44	.51	3.64		
Ideal: Board Presidents					
Non-performance factors	44	.58	2.93	-2.22	.0317
Performance factors	44	.90	3.23		
Ideal: Board Presidents					
Non-performance factors	44	.49	3.07	-6.05	.0001
Performance factors	44	.46	3.68		

Given these results, it was determined that superintendents didn't differ in rating the relative importance of current practice performance-based compensation when compared to non-performance based compensation factors for superintendents. However, in the ideal situation, superintendents did rate performance-based compensation factors higher than non-performance based factors with a higher aggregate response.

Tests of significance were also applied to aggregate responses of board presidents. Board presidents differed somewhat by placing more importance on performance-based factors than non-performance based factors in both the current situation and in the ideal situation. The ideal difference was dramatic and strong in favor of using performance-based factors given the opportunity in an ideal or preferred situation.

Other tests involved comparing aggregate means for superintendents and board presidents between current practice and ideal situations in non-performance and performance-based areas. The results of these analyses indicated significant differences between current practice and ideal situations on non-performance and performance factors for superintendents, but not for board presidents (Behounek, 1996).

Another analytical exercise was undertaken to determine if the perceptions about performance-based compensation had any relationship to the perceived quality assessment instrument data. Superintendents perceptions on performance-based compensation were categorized in three groups " low support, moderate support, and high support " relative to strength. These groups then were analyzed with ANOVA to determine if there is a difference related to the superintendents' perceptions. Low support was defined as a mean score of less than 3.0, moderate included means between 3.1 and 4.25, and high included mean scores greater than 4.26. The categories were selected for purposes of comparison, and a confidence level of .05 was used. The results of the analysis are presented in the following Table.

Table 16. ANOVA summary table comparing Quality Effectiveness Index rating of school systems by perceived support of performance-based compensation by superintendents

Source	**df**	**SS**	X^2	*F* **ratio**	*FP*
Between groups	2	.21	.1	1.23	.303
Within groups	41	3.43	.08		
Total	43	3.64			

Group	*N*	*X*	**St. Dev.**
Low support for performance compensation	6	3.42	.22
Moderate support for performance compensation	35	3.35	.30
High support for performance compensation	3	3.10	.34
Total	44	3.34	.29

Statistically, there is no significant difference in the rated quality of the school systems based upon the level of the superintendent's support for performance-based compensation. The results, given full awareness of the small sample size, do display some interesting configurations, and a trend is somewhat evident. The group of superintendents with the highest mean response indicating high support for performance-based compensation, served in districts with the lowest mean quality index (3.10). The group of superintendents with the lowest mean response indicating low support for performance-based compensation, served in districts with the highest mean quality effectiveness index (3.42). As school systems are categorized by their superintendent's support of performance-based compensation, the aggregate mean scores of the quality of the districts increase as the support decreases.

To test the significance of this perceived relationship, a correlation (Pearson product-moment) was calculated to determine if the strength of the superintendents' support of performance-based compensation was related to the ratings of perceived system quality reflected in the aggregate quality effectiveness index ratio. The results of the analysis are shown in Table 17 below.

Table 17. Correlation of school system quality effectiveness index by level of superintendents' support for performance-based compensation

N	**Covariance**	*R*	R^2
44	-.03	-.18	.03

The results in Table 17 reflect a slight negative correlation between these two variables. A negative correlation of -.18 is determined to be "little if any" (Hinkle, 1988). Weak as it is, the relationship between perceived organizational quality and support for performance-based compensation places imperfect credibility on Deming's theory that performance pay is detrimental to an organization (Deming, 1986).

Analysis of Financial Factors and PQAI

The impact of financial soundness upon organizational quality has been demonstrated in a number of ways in the private sector, but has not been demonstrated unequivocally in public education (Kirchoff, 1996). Despite a paucity of research demonstrating the influence of financial wealth upon educational institutional quality, there has been a belief that financial stability of a system may contribute to collaboration efficacy and to constancy of purpose (Schmoker and Wilson, 1993).

Several financial characteristics of educational institutions were identified and assembled and considered in relationship to the quality effectiveness index generated by the Perceived Quality Assessment Instrument. The factors included the financial solvency ratio (FSR), revenues per pupil (R/P), expenditures per pupil E/P), undesignated unreserved fund balance per pupil (UUFB/P), unspent balance per pupil (UB/P), and the transportation cost per pupil (TC/P). These six factors were examined in two ways, first with a correlational matrix, using Pearson product-moment techniques, and then with the t-test (two-tailed) between the high and low financial factors compared to the PQAI score. In other words, the most financially sound districts and the least financially sound districts were compared as to the level of PQAI score, or perceived quality effectiveness index.

To obtain the highest and lowest categories, the sample group of school systems was divided in to three equal groups of fourteen school districts each (this analysis involved the use of 42 of the 44 school systems) based upon the districts perceived quality effectiveness index. The index is the ratio of the perceived current mean score on the PQAI to the perceived ideal mean score on the PQAI for the school district. The means and standard deviations for the sample group, the high PQAI group, and the low PQAI group are shown in the Table below.

Table 18. Statistical data for the PQAI sample, high, and low groups

Group	*N*	Mean	St. Dev.
Sample Total PQAI	42	.796	.080
High PQAI Group	14	.885	.044
Low PQAI Group	14	.716	.047

Correlations were calculated for each of the financial characteristics between the high quality effectiveness index group and the low quality effectiveness index group. The results of those calculations are illustrated in Table 19 below.

Table 19. Correlations between financial characteristics and PQAI

	Financial characteristics					
PQAI Factor	**FSR**	**R/P**	**E/P**	**UUFB/P**	**UB/P**	**TC/P**
Sample Group Total	-0.011	0.103	0.113	-0.009	0.099	0.226
Leadership	-0.083	0.144	0.189	-0.077	-0.028	0.115
Inf. and analysis	-0.037	0.095	0.095	-0.031	0.111	0.326*
Strategic planning	0.017	-0.002	-0.007	0.006	0.124	0.079
HRD and mgmt.	-0.104	0.240	0.269	-0.060	0.193	0.238
Mgmt. of process quality	-0.091	0.069	0.056	-0.092	0.086	0.177
Quality and op. results	0.024	0.097	0.106	0.029	0.061	0.216
Client focus and satisfaction	0.012	0.120	0.128	0.112	0.112	0.259

*$p < .01$

A statistically significant relationship was shown between the transportation cost per pupil and the PQAI of the sample in the information and analysis category of the perceived quality assessment instrument, however, the variance (r2) was calculated to be only 0.106. This means that only 10.6% of the variance in the transportation cost per pupil was attributable to the PQAI and that almost 90% of the variance was attributable to other factors.

T-tests were also calculated for each of the financial factors comparing the high Quality Effectiveness Index group with the low Quality Effectiveness Index group to determine if the differences between the groups on any financial factors were significant. The results of this analysis is shown in the table below.

Table 20. *T* tests between the high and low PQAI groups and financial characteristics

Group	**PQAI**	**FSR**	**R/P**	**E/P**	**UUFB/P**	**UB/P**	**TC/P**
Sample Group Total	9.706*	-0.441	0.352	0.379	-0.440	0.411	1.016
Leadership	8.878*	-0.092	2.336*	0.760	0.063	0.763	-0.014
Inf. and analysis	10.665*	-0.551	0.307	0.171	-0.486	0.185	2.250*
Strategic planning	7.976*	-0.262	-0.363	-0.527	-0.309	1.044	0.000
HRD and mgmt.	6.136*	-0.216	1.758	1.581	-0.042	0.378	1.493
Mgmt. of process quality	10.367*	-0.353	0.221	0.269	-0.513	0.568	0.572
Quality and op. results	8.146*	0.218	0.901	1.095	0.319	-0.130	1.751
Client focus and satisfaction	11.159*	0.661	1.077	1.124	0.746	0.897	1.728

*$p < .0001$

Significant differences were found between the high and low group PQAI means, for the revenues per pupil in the leadership category, and the transportation cost per pupil in the information and analysis category. One of the significant t-test results was in the area of leadership and revenues per pupil. Another was in the area of transportation costs per pupil and the information and analysis sub-scale. While statistical significance was found between two of the financial characteristics, forty-six differences were not found to be significant.

Analysis of the Perceived Quality Assessment Instrument

The study also sought to determine if the PQAI items on the perception of quality scales of current and ideal status would align consistently with the seven a-priori determined dimensions based on the PQAI (Baldrige) criteria. A factor analysis was used based on a varimax rotation technique. The results of the analysis revealed three possible factors with Eigenvalues greater than 1.0 on both the current and ideal scales.

The three factors accounted for 42.3% and 43.3% of the total variance on the current and ideal scales respectively. The distribution of the number of items from each PQAI category among the empirical factors suggested is shown in Table 21. For example, the six a-priori items for the Leadership category (current responses) were all contained in factor 3; however the eight items for the Customer Focus and Satisfaction were distributed among factors 2, 3, 6, and 7.

Table 21. Comparison of the current and ideal a-priori PQAI dimensions with empirical factors

Scale	Items	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8
Eigenvalue		16.28	1.50	1.27	.99	.79	.73	.59	.49
Current									
Leadership	1-6			6					
Information & Analysis	7-12		1		5				
Strategic Quality Planning	13-18	3				5			
H.R. Devel. & Mgt.	19-25	5				2			
Mgt. of Process Quality	26-31	2	3						
Quality & Op. Results	32-37		5						
Customer Focus	38-45		3	1			1	3	
Eigenvalue		16.33	1.99	1.17	.89	.68	.56		
Ideal									
Leadership	1-6				6				
Information & Analysis	7-12			1		4	1		
Strategic Quality Planning	13-18		6						
H.R. Devel. & Mgt.	19-25		1			6			
Mgt. of Process Quality	26-31	5	1						
Quality & Op. Results	32-37	5					1		
Customer Focus	38-45	7							

The results, as shown in the Table above, suggest that the seven original a-priori dimensions may be reduced to three factors on both scales. However, the majority of the items in the Information and Analysis dimension load on factor 4 on the current scale and factor 5 on the ideal scale and the majority of items in the Human Resource Development and Management load on factor 4 in the ideal scale, indicating the importance of these additional factors.

On the ideal scale, three dimensions (Management of Process Quality, Quality and Operational Results, and Customer Focus and Satisfaction) seem to load on one factor (factor 1) which suggests that respondents were not able to differentiate among these three dimensions as originally conceptualized or that the three were perceived as measuring the same thing. The factor analysis results show the items grouped differently than on the original instrument showing that the instrument may be measuring different factors than originally conceived and the items may need to be regrouped. However, when combined with the results of the correlation matrix, it appears that, overall, the instrument is measuring one underlying concept on both scales.

Summary and Conclusions

This paper presents the findings of studies of public schools in Iowa in the area of perceived quality assessment conducted at Iowa State University. Demographic characteristics of the respondents on the Perceived Quality Assessment Instrument from forty-four school districts were described by position, home annual income, gender, age, level of education, and years experience in current or a similar job.

The research project undertaken incorporated several studies of quality improvement characteristics of public schools. The project resulted in a compendium of coordinated research aimed at learning more about the relationships and effects of quality improvement efforts with other factors of school district operations.

Several dissertations resulted from the studies, including writings by Rashid M. Bax, Pamela D. Johnson, Thomas J. Behounek, and Joseph E. Kirchoff, over a three-year period. All of the dissertations were supervised by William K. Poston Jr. at Iowa State University during the period 1993-1995. Each of the study components (parts of the total study) gathered specific data relevant to the issues of the individual study component. The individual study components (doctoral dissertations) focused on the following issues:

- Assessment of quality improvement climate in community colleges (Bax, 1994).
- Teachers' perceptions of training programs and their relationships to total district perceptions of quality management (Johnson, 1995).
- Performance-based pay of chief executive officers and effects upon quality improvement processes in school organizations (Behounek, 1996).
- Financial characteristics of school organizations and relationships to quality management factors (Kirchoff, 1996).

Continuous quality improvement is a recent phenomenon in educational institutions, but it has been popular in business and industry since the late 1970's in the United States. The management philosophy inherent in continuous quality improvement has been largely based upon the ideas of W. Edwards Deming. The incorporation of Deming's ideas into educational organizational thinking has been gradual and focused on application of the quality concepts into everyday practice. Given the goals of public education, questions arise as to whether or not the use of continuous quality improvement principles have relevance in the public school organizational environment. Moreover, it is even more important to know what relationships and effects result from the application of quality improvement strategies in the public school setting.

This series of studies conducted over a period beginning in 1993 and culminating in 1996, was designed to assess the perceptions of school district stakeholders about the quality of their school district in terms of the seven quality dimensions of the Baldrige Award criteria. The

Baldrige criteria were used in the development of an instrument that was employed in evaluating the levels of quality in public school organizations. This version of the Baldrige instrumentation was developed and modified for use in the studies (Poston and Bax, 1994). The instrument, in two versions, was administered to 44 independent local school systems and to 15 community college in the state of Iowa. and focused on the following Baldrige Award areas:

1. Leadership
2. Information and Analysis
3. Strategic Quality Planning
4. Human Resource Development and Management
5. Management of Process Quality
6. Quality and Operational Results
7. Client Focus and Satisfaction

The criteria for the instrument were modified to fit the public school environment, and the instrument contained 45 items (see Appendix). Validity and reliability of the instrument was established using statistical inquiry, panel review, and research validation procedures. The instrument was found to be both valid and reliable for use in educational organizations and institutions. Respondents were instructed to indicate the current state of affairs on each item of quality, as well as the ideal, or preferred, state of affairs for each item.

Two other instruments were developed for use as a part of the studies. Both instruments were developed based upon information from professional literature, and verified by a panel of reviewers. In the study component by Pamela Johnson, the Staff Development Questionnaire was developed and employed, and the Executive Compensation Questionnaire was developed and used in the study component managed by Thomas Behounek. Joseph Kirchoff's study employed the use of publicly-documented data available from state government agencies.

Study Procedures

In the comprehensive studies, it was important to adequately select a population of school districts reflective of the State of Iowa. The population consisted of forty-four (44) school districts selected, invited, and willing to participate in the studies voluntarily. The districts represented more than ten percent (10%) of the districts in the State of Iowa, and they were generally reflective of the state in terms of size, economic conditions, and geographic distribution. The chief executive officers of each school system were requested to assign random distribution of the instrumentation to the stakeholders in their districts to a third party, and respondents were asked to complete all instruments, following the directions provided.

Those eligible for completing the Perceived Quality Assessment Instrument were the superintendents, all board members, two members of the administrative staff, three members of the support staff, five teachers, and two high school students. The comprised a total of 18-20 persons from each school district, depending upon the size of the board (five to seven members). All responses to all instruments were anonymous and voluntary. The total possible number of questionnaires to be returned amounted to 720 each, of which 471 instruments were received. This amounted to a return rate of approximately 65 percent. Four districts were eliminated from the final analysis of data due to insufficient returns from those districts.

The Staff Development Questionnaire was planned to be completed by six teachers in each of the forty-four districts, and they were not allowed to be the same teachers used in the PQAI instrumentation response. Of a possible 264 instruments, 196 were returned for a response rate of approximately 80 percent.

The Executive Compensation Questionnaire was administered to all superintendents and board presidents of the 44 school districts. Remarkably, all 44 superintendents and all 44 board presidents responded by completing and returning the questionnaire.

The financial information used in the studies was obtained from officially-filed audit reports from each school district from the offices of the Iowa State Department of Education.

All data were gathered by June, 1994, and all responses were entered into a spreadsheet program for analysis and computation of means for current and ideal status. The quality effectiveness index was also computed for each district on each item, sub-area, and for the total instrument. District means were also computed for all instrument data from the Staff Development Questionnaires and the Executive Compensation Questionnaire. Districts were then rank-ordered from high to low, and different groupings were used for data analysis and comparisons.

Findings

Of those responding on the PQAI, superintendents (30 each) comprised 6.6% of the sample group, teachers (132) comprised 28.8%, support staff (76) comprised 16.6%,, board members (108) made up 23.6%, and students (44) comprised 9.6%. Demographic data were also gathered on each of the responding individuals, including occupational positions, income, gender, age, levels of education, and work experience of individual respondents.

Differences were noted between male and female respondents, which was significant overall, but not in any of the separate sub-areas of the instrument when broken down by gender. Teachers were noted to differ from board members, and differences between administrators and teachers were also noted on the PQAI. Respondents with a baccalaureate degree also differed from those without the degree. No significant differences were noted in the demographic categories of age, income levels, or job experience.

PQAI Findings

The perceived current level of quality and the ideal level of quality were calculated, and the means for the current status ranging from 2.81 to 3.92. Ideal means ranged from 3.79 to 4.66. Quality effectiveness indices were calculated, and ranged from .6532 to .9506. Two districts had a QEI above .90, and three had a QEI below .70. The majority (28 or 70%) of districts had a QEI above .75 indicating that those districts perceive that they are achieving at least 75% of their quality management goals.

Interrelationships among the seven sub-areas of the PQAI were also found using a matrix of relationships, with moderate to high correlations among the factors. A correlation of .77 was found between human resource development and management and strategic quality planning, .74 between client focus and satisfaction and quality and operational results, .73 for human resource development and management and management of process quality, and .71 between management of process quality and operational results.

A determination was also sought to establish if the PQAI differentiated in terms of quality between high ranking and low ranking school districts. Inferential statistics were used to assess the difference and established a significant difference between the ten highest rated and the ten lowest rated groups. There was not only a significant difference between the high and low groups' quality effectiveness index, but there were also significant differences between the groups in each of the seven dimensions or sub-areas of the PQAI instrument. Finally, the Perceived Quality Assessment Instrument was evaluated utilizing factor-analysis techniques, and results indicated that some items may need to be regrouped into three categories or sub-areas empirically developed but undefined.

Staff Development Findings

A significant positive relationship was found between the perceived quality of district staff development and the perceived quality effectiveness index of the districts. District means

on the Staff Development Questionnaire were used to categorize the districts into four equally-sized groups according to their ranks, and there were significant differences between all four groups in their rankings of staff development practices. In other words, significant differences were found between the districts ranked highest and lowest in perceived staff development quality on the quality effectiveness index. The top ten districts demonstrated a modest correlation of .564 between the means of the QEI and the SDQ, indicating that districts which have higher quality staff development programs also have higher quality overall district management as measured by the Quality
Effectiveness Index slightly more than half of the time.

Significant relationships were also found between ratings of perceived levels of district staff development and current ratings on three of the seven dimensions of the Perceived Quality Assessment Instrument. Client focus and satisfaction, quality and operational results, and management of process quality were found to correlate positively and have predictive value with perceptions of quality in staff development. The other four areas -- leadership, information and analysis, strategic quality planning, and human resource development and management -- showed limited correlative or predictive value.

Executive Compensation Findings

Board presidents and superintendents showed agreement that the student enrollment levels should have little effect in determining compensation levels for the superintendent, and also agreed that standardized testing should not be used for such purposes. That teacher union increases should be considered in any compensation increases for superintendents was also an area of agreement, as was the area calling for use of job description factors in compensation decisions.

The differences between performance-related and situational-related (nonperformance) factors were evidenced between board presidents and superintendents with the former more supportive of using performance-based criteria than the latter. However in an ideal setting, superintendents were more inclined to consider performance-based criteria more important than currently. Performance-based compensation initiatives would find little or no support for such practices from these studies. Ironically, the use of feedback and assessment information for improvement efforts appears to be contradicted to a degree in the perceptions of key organizational leaders -- superintendents and board presidents.

A very weak finding was that there may be some credence to W. Edwards Deming's proscription of performance-based (merit) pay, in that a weak inverse relationship was found between performance-based criteria compensation support and the perceived quality of the systems. Deming's notion that performance-based pay might have a deleterious effect upon organizational quality was confirmed in this study, albeit tenuously.

Financial Factors Findings

Five selected financial characteristics and the financial solvency ratio were studied in relationship to perceived quality of the schools in the study. The five factors included revenues per pupil, expenditures per pupil, undesignated, unreserved fund balance per pupil, unspent balance per pupil, and transportation cost per pupil. Means of the financial characteristics for the study population were calculated, and statistical analyses were completed.

Significant differences, using t-tests, were found between the high and low QEI groups in two areas -- revenues per pupil in the leadership sub-area, and transportation cost per pupil in the information and analysis category. The other 46 possibilities were not found to be significant. The perceived leadership quality increases slightly as the amount of revenue per pupil increases, and vice-versa for transportation cost and perceived quality in information and analysis. No correlation was found between the sample schools financial characteristics and their any PQAI rating area with one exception -- transportation cost per pupil and information and analysis.

Conclusions and Recommendations

The comprehensive studies of quality improvement perceptions and other selected factors in Iowa schools revealed several encouraging results and disappointments simultaneously. The Perceived Quality Assessment Instrument appeared to function relatively well as evidenced by its significant discrimination between high and low rated systems. It also discriminated between high and low perceived quality school districts on the sub-areas of the instrument. Reliability and validity of the instrumentation was well within tolerable limits, and the ease of administration of the instrument contributed to a reasonable return of instruments.

The instrument actually seems to be measuring three empirically-determined factors, but these remain undefined pending further research. The strength of the instrument's utility lies in its capability to accurately measure a school district's quality improvement status against a definable, solid standard, but currently, the instrument is in need of further refinement. The quality effectiveness index or ratio provides a convenient, easily understood criterion for comparing districts against the fixed standard, and when used, it does reveal to the organization its level of quality attainment *against its own definition of ideal quality*.

The efficacy of staff development programs in relationship to organizational quality was supported in the results of these studies, although the relationship between staff development and overall quality was not as strong as might be expected. However, effective staff development calls for resources, time, and strong leadership, and teachers' ratings may not be aware of the demands upon those things which require choices and limitations.

An encouraging finding was the relationship of staff development quality and the perceived organizational quality levels of the PQAI. It would stand to reason that districts that make strong commitments to staff development would also make commitments to other areas of quality improvement as well. In addition, training of staff should have some beneficial effect upon effective practice and organizational functioning. Excellence in training reasonably should be expected to produce excellence in organizational behavior.

Individual analyses of all items in the seven dimensions of the PQAI show the existence of moderate to strong relationships among the group, indicating that they are dealing with factors that are connected. The means of the ten districts ranking highest and lowest in staff development quality were compared with their means on the Quality Effectiveness Index, and the top ten were significantly different than the lowest ten, suggesting that although quality staff development is one of the factors important in the implementation of quality management.

Another important outcome of these studies was found in both the executive compensation and financial factors components -- that some things are not related to quality. The lack of strong confirmation that performance-based pay is related to or contributory to some level of difference is an important finding in light of continuing efforts to tie job compensation to individual performance. Such a reward-penalty approach was not supported in these studies, but was not refuted all together. The notion of performance-based or merit pay systems will no doubt continue to be popular in certain political arenas, but it did not gain vindication in these studies.

Many educators have sought to establish a clear, straight line relationship between financial wealth of schools and the level of quality within the system. Again, little support was brought about in these studies for such an inclination. Although a couple of relationships were found between financial characteristics and perceived quality in these studies, but the absence of such verification encourages researchers and educational practitioners to seek contributory factors in other directions.

Recommendations

Based on the findings and conclusions of these studies, a number of recommendations emerge. The Perceived Quality Assessment Instrument appears to have considerable merit as a tool in determining the levels of quality in school systems, and as a tool, it might be effectively used as a team informational device in planning and developing improvement efforts. The seven dimensions provide a framework for improvement actions at all levels in school organizations. There should be no hesitation to proceed with continued development, research, and refinement of the perceived quality assessment process.

Greater use of the instrument in research activities could fashion improvement in the instrument's quality, given feedback and revision of items and organization over time. The three empirically-derived measurement areas provide an opportunity for other developmental activity.

In addition, there is considerable room for use of the instrumentation in other studies and research efforts. The PQAI and similar tools could focus on the improvement of practices and programs in schools by making results-oriented information available. The common understanding of quality is yet to be defined in educational organizations, and any instrument like the PQAI should be considered for use in improving school performance, planning, training, and institutional assessment. Without better understandings of these and other factors, improved productivity in school organizations rests upon a better grasp of quality than currently exists.

Quality improvement work should also involve the use of growth-monitoring strategies. Using the PQAI as a "baseline" establishing device after further refinement and development, would enable districts to measure current status, establish directions for improvement, and monitor progress toward clearly defined ends and goals. Trend data, establishment of performance levels, and benchmarking information can all be derived from use of the PQAI, and more effective evaluation of system progress becomes attainable.

Closing Comments

The purpose of these studies centered on the determination of whether or not the Perceived Quality Assessment was effective in measuring factors commonly associated with excellence, efficacy, and effectiveness in school organizations. A second purpose was to identify the extent of any relationships and the nature of any associations among variables measured by the PQAI and other measures, and the final purpose was to find whether or not operations and management actions in schools have an impact or influence upon the perceived level of quality within the system.

Given the findings and outcomes of these studies, the studies' purposes have been achieved. However, that is only part of the voyage. Organizational quality is a continual process, not a destination, and enduring perseverance toward finding new and better ways to improve overall school effectiveness, use of resources and capabilities, and the overall development and well-being of clientele is essential for human progress. The journey must go on.

References

Bates, M. and Stachowski, E. (1991). Planning comprehensive staff development programs. San Francisco, CA: Association for Supervision and Curriculum Development.

Bax, R. (1994). An assessment of the quality improvement climate as perceived by community college leadership in Iowa. Doctoral dissertation, Iowa State University, Ames, Iowa.

Bax, R. M. (1994). Relationships between quality ratings, job satisfaction, and organizational commitment in school districts implementing total quality management. (Unpublished paper, Iowa State University, Ames, Iowa, 2 May 1994).

Behounek, T. J. (1996). A study of the perceived quality factors and methods of awarding salary increases for superintendents in selected school districts. Doctoral dissertation, Iowa State University, Ames, Iowa.

Borg, W.R. and Gail, M.D. (1989) Educational research: an introduction. New York: Longman.

Bradley, L.H. (1993). Total quality management for schools. Lancaster, Pennsylvania: Technomic Publishing Co. Brown, E.H. (1992). On Deming and school quality: a conversation with Enid Brown (in Brandt, R., 1992). Educational Leadership, 50(3), 28-31.

Deming, W.E. (1986). Out of the crisis. Cambridge, MA: Massachusetts Institute of Technology Center for Advanced Engineering Study.

Deming, W.E. (1990). Quoted in Walton, M. Deming Management At Work. New York, NY: Putnam. P. 11.

Drystad, J. (1994) Quoted in Rubach, L. Transforming U.S. education with the quality movement. Quality Progress. 27(9), 93-95.

Frase, L.E., English, F.W., and Poston, W.K. (1995). The curriculum management audit: improving school quality. Lancaster, PA: Technomic Publishers.

Fullan, M.G. (1991). The new meaning of educational change. New York: Teachers College Press.

Graff, L. (1995). Committing to growth through a learning organization. (Paper published on the Horizon "Listserv" online newsletter, 15 November 1995). Horizon@gibbs.oit.unc.edu.

Guskey, T.R. and Sparks, D. (1991). What to consider when evaluating staff development. Educational Leadership, 47(3), 73-76.

Guskey, T.R. (1994). The most significant advances in the field of staff development over the last twenty-five years. Journal of Staff Development, 15(4), 5-6.

Hinkle, D.E., Wiersma, W., and Jurs, S.G. (1988). Applied statistics for the behavioral sciences. (2nd Ed.) Boston: Houghton Mifflin.

Johnson, P. D. (1995). An examination of teachers' perceptions of district staff development programs and their relationship to perceptions of district quality management. Doctoral dissertation, Iowa State University, Ames, Iowa.

Joyce, B. & Showers, B. (1988). Student achievement through staff development. New York: Longman.

Kirchoff, J. E. (1996). A study of quality improvement and its relationship to financial characteristics of selected Iowa public school districts. (Doctoral dissertation, Iowa State University, Ames. Iowa, 1996).

Knudson, B. (1993). Continuous improvement profile. Minneapolis, MN: Positive Directions, Inc.

Lezotte, L. (1994). The nexus of instructional leadership and effective schools. The School Administrator, 51 (6), 20-23.

Locke, E.A. (1976). The nature and cause of job satisfaction. (in Dunnette, M.D., ed.). Handbook of Industrial and Organizational Psychology. Chicago, Illinois: Rand McNally.

Marchese, T. (1993). A time for ideas. Change, May-June, 1993, P. 11.

McBride, R., Reed, J., and Dollar, J. (1994). Teacher attitudes toward staff development: a symbolic relationship at best. Journal of Staff Development, 15(2), 36-41.

Nunnally and Durham, 1975 (see Behounek)

Orlich, D.C. (1983). Some consideration for effective inservice education. The Clearing House, 56(5), 197-201.

Porter, L., Steers, R. Mowday, R., and Boulian, P. (1974). Organizational commitment, job satisfaction, and turnover among psychiatric technicians. Journal of Applied Psychology, 59, 603-609.

Poston, W.K., Jr. and Frase, L.E. (1991). Local teacher alternative compensation programs: rolling boulders up the mountain of reform. The Phi Delta KAPPAN, 17(4) 317-320.

Poston, W. K. Jr. (1995). Making Governance Work: Total Quality Education for School Boards. Newbury Park, California: Corwin Press. Poston, W. K. Jr. (1995). Studies of the aspects of the use of the Malcolm Baldrige national quality award criteria in the assessment of educational organizations. (Paper presented to the National Conference on Education at the American Association of School Administrators, New Orleans, Louisiana, February, 1995).

Poston, W.K. Jr. and Bax, Rashid, PQAI

Rebore, R. (1991). Personnel administration in education: a management approach. Englewood Cliffs, NJ: Prentice Hall.

Rowe, J. (1994). Outmoded leadership ideas won't work in (the) 21st Century. (Article in the Des Moines Register, 27 June 1994, Page 8-B).

Schenkat, R. (1993). Deming's quality: our last but best hope. Educational Leadership, 51(1), 64-65.

Schenkat, R. (1993). Quality Connections: Transforming Schools Through Total quality Management. Alexandria, Virginia: American Association of School Administrators.

Schmoker, M. and Wilson, R.B. (1993). Transforming school through total quality education. The Phi Delta KAPPAN, 74(5), 389-395.

Senge, P. (1990). The fifth discipline: the art and practice of the learning organization. New York: Doubleday.

Showers, B. (1985). Teachers coaching teachers. Educational Leadership, 42(7), 43-48.

Sparks, D. and Loucks-Horsley, S. (1989). Five models of staff development for teachers. Journal of Staff Development, 10(4), 40-57.

Sparks, D. and Vaughn, S. (1994). What every school board members should know about staff development. Journal of Staff Development, 15(2), 20-22.

Teeter, D.J. and Lozier, G.G. (1993). Six foundations for total quality management. In D.J. Teeter and G.G. Tozier, (eds.). Pursuit of quality in higher education: case studies in total quality management. New Directions for Institutional Research, No. 78, pp. 5-11. San Francisco, California: Jossey-Bass Publishers.

Teigland, M.D. (1995). A study of the beliefs for total quality management comparing superintendents, board members, and classroom teachers in Iowa schools. Doctoral dissertation, Iowa State University, 1993.

Templeman, T.P. and Peters, J. (1992). Teaching research inservice model. Monmouth, OR: Western Oregon State College.

U. S. Department of Commerce. (1995). The Malcolm Baldrige National Quality Award: Education Pilot Criteria, 1995. National Institute of Standards and Technology, Washington, DC.

U.S. General Accounting Office. (1993). System-wide education reform: Federal leadership district-level efforts. Report to Congressional Requesters. U.S.G.A.O.: Washington, DC.

Vincoli, J. (1991). Total quality management and the safety and health professional. Professional Safety, 6, 27-32.

Wade, R.K. (1985). What makes a difference in in-service teacher education? A meta-analysis of research. Educational Leadership, 42(4), 48-54.

Wood, F., Killian, J., McQuarrie, F., and Thompson, S. (1993). How to organize a school based staff development program. Alexandria, VA: Association for Supervision and Curriculum Development.

Appendix:

Sample Instruments

Perceived Quality Assessment Instrument

Staff Development Questionnaire

Executive Compensation Questionnaire

Note: Copies of instruments may be obtained from:
William K. Poston Jr.
Department of Professional Studies

N229 Lagomarcino Hall
Iowa State University
Ames, Iowa 50011

About the Author

William K. Poston Jr., Ed.D.

Associate Professor, Educational Administration
Department of Professional Studies
N225 Lagomarcino Hall
Ames, Iowa 50021
Phone: (515) 294-9468
Fax: (515) 294-4942
Email: wkpjr@iastate.edu
Iowa State University

Dr. Poston is currently Associate Professor of Educational Administration in the Department of Professional Studies at Iowa State University in Ames, Iowa. Bill is a former math and physical science teacher and an experienced administrator, with 25 years of experience in school administration including 15 years as a superintendent in the Flowing Wells (Tucson) and Kyrene (Tempe-Phoenix) Schools in Arizona and in the Billings Public Schools, in Montana. Bill earned his B. A. at the University of Northern Iowa, and his Ed. S. and Ed. D. at Arizona State University. He has many distinctive professional achievements, including service as the youngest-elected international president of Phi Delta Kappa, and selection as an Outstanding Young Leader in American Education in 1980.

Bill has authored many articles and has recently published four books: Making Schools Work: Practical Management of School Operations (Corwin Press, 1992), Effective School Board Governance (Phi Delta Kappa International, 1994), Making Governance Work: Total Quality Education for School Boards, (Corwin Press, 1994), The Curriculum Audit: Improving School Quality (Co-author, Technomic Publishing, 1995).

Dr. Poston teaches courses in contemporary management, strategic planning, school finance, and educational leadership at Iowa State University. He is also the executive director of the Iowa Superintendents' Academy and the Iowa School Business Management Academy. He continues to provide extensive service to schools in the areas of evaluation, curriculum management auditing, performance-based budgeting, and continual quality improvement.

Bill has been married for 35 years to his wife, Marcia, and they have two daughters, both of whom are teachers in the West Des Moines Community Schools.

Copyright 1997 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://olam.ed.asu.edu/epaa

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, Glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692)

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

John Covaleskie
Northern Michigan University

Andrew Coulson
a_coulson@msn.com

Alan Davis
University of Colorado, Denver

Sherman Dorn
University of South Florida

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
hmwkhelp@scott.net

Thomas F. Green
Syracuse University

Alison I. Griffith
York University

Arlen Gullickson
Western Michigan University

Ernest R. House
University of Colorado

Aimee Howley
Marshall University

Craig B. Howley
Appalachia Educational Laboratory

William Hunter
University of Calgary

Richard M. Jaeger
University of North Carolina--Greensboro

Daniel Kallós
Umeå University

Benjamin Levin
University of Manitoba

Thomas Mauhs-Pugh
Green Mountain College

Dewayne Matthews
Western Interstate Commission for Higher Education

William McInerney
Purdue University

Mary P. McKeown
Arizona Board of Regents

Les McLean
University of Toronto

Susan Bobbitt Nolen
University of Washington

Anne L. Pemberton
apembert@pen.k12.va.us

Hugh G. Petrie
SUNY Buffalo

Richard C. Richardson
Arizona State University

Anthony G. Rud Jr.
Purdue University

Dennis Sayers
University of California at Davis

Jay D. Scribner
University of Texas at Austin

Michael Scriven
scriven@aol.com

Robert E. Stake
University of Illinois--UC

Robert Stonehill
U.S. Department of Education

Robert T. Stout
Arizona State University

---

archives | abstracts | editors | board | submit | comment | subscribe | search

---

This article has been retrieved 1535 times since November 24, 1997.

Education Policy Analysis Archives

Volume 5 Number 20 November 24, 1997 ISSN 1068-2341

A peer-reviewed scholarly electronic journal.
Editor: Gene V Glass Glass@ASU.EDU.
College of Education
Arizona State University,Tempe AZ 85287-2411

Copyright 1997, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

The Use and Abuse of Socrates in Present Day Teaching

Anthony G. Rud Jr.
Purdue University

Abstract The Greek philosopher Socrates is used as an example of a master teacher in in many contexts, from elementary school discussions, to college philosophy classes, to law school. I examine a number of current uses of Socratic teaching, and expose inconsistencies among them. I analyze critically recent practitioners of Socratic teaching, such as Mortimer Adler, and I consider how the celebrated primary teacher Vivian Gussin Paley enacts the Socratic legacy in a novel way. I argue that the misuse, or abuse, of the Socratic legacy occurs chiefly when his teaching is interpreted narrowly as a pedagogical technique devoid of context and irony.

Introduction

The title of my paper is a deliberate play upon Friedrich Nietzsche's well-known essay, The Use and Abuse of History (1874, 1979). In that work, Nietzsche turned his eye upon his culture to decry what he termed its "malignant historical fever" (p. 4). He believed that a mere studying of the past, particularly by self-absorbed scholars, was not a vital use of historical tradition. Rather, knowledge of the past must instead serve both the present and future (p. 22), and not become merely an abstract item devoid of the context that initially gave it life (pp. 11-12).

Today, such a figure from the past serves as an important model and inspiration for much current pedagogy. The Greek philosopher Socrates is used as an example of the master teacher in many contexts, from philosophy classes to law school. There is effort underway to incorporate "Socratic" dialogue into many programs at the precollegiate level (Lipman et. al., 1980; Obermiller, 1989). On the surface, then, it would seem that this particular bit of history, brought to life for us through Plato, Xenophon, and Aristophanes, is also alive in teaching and learning, beyond the careful scholarship of the university classicist or philosopher.

This diversity in the use of the legacy of Socrates in current pedagogy does signify a vital tradition. Many of the uses of this legacy are admirable. Yet, understandings of a "Socratic method" differ widely. There is, for example, disagreement over whether Socrates offered a pedagogical method as that term is understood today. I propose to examine a number of uses of Socratic pedagogy in different contexts in order to show inconsistency among them, particularly in reference to the Platonic Socrates. I build upon other recent work (Haroutunian-Gordon, 1991; Burbules, 1993; Pekarsky, 1994) that explore the Socratic legacy for education, while offering insights into additional recent Socratic practitioners (Paley, 1986, 1990; Adler, 1982, 1990; Weiss, 1987).

I conclude that there is widespread use of the term "Socratic" in descriptions of certain types of teaching. Yet, when Socratic teaching is taken to mean everything from dialectical examination of philosophical issues of justice, the good, and the like, (Gray, 1988) to the use of questions by a teacher, independent of the subject matter (Kay and Young, 1986), there needs to be a clearer understanding of the uses of the Socratic legacy for teaching.

Background: Some Recent Commentators on Socrates

Several current critical views of Socrates should be taken into account if we are to fully understand and be able to appraise critically that legacy. Moreover, such criticism is a key element in a determination of the uses of Socrates for present day teaching, as I shall argue at the end of this paper. Recent commentators on both the historical and Platonic Socrates, among them Bruce Kimball (1986), I. F. Stone (1988), and Friedrich Nietzsche (1872, 1956), have been critical of Socrates and his legacy. Stone sees a Socrates as the enemy of the nascent Greek democracy, Nietzsche portrays a degenerate destroyer of the heroic legacy of the tragic age of Greece, and Kimball rues the emphasis of Socratic rationality over Ciceronian oratory in liberal education.

Nietzsche, though of at least two minds about Socrates (Dannhauser, 1974, especially pp. 269f.), began his career with a full-force attack upon the Greek. In The Birth of Tragedy (1872, 1956), he laments the emergence of the Socratic spirit of exhaustive analysis that put an end to the Apollinian-Dionysian mix that spawned early Greek tragedy. Nietzsche notes that Socrates was incapable of appreciating the earlier tragedians, like Aeschylus, and only attended the plays of Euripides (whom Nietzsche sneeringly calls the first rational tragedian (pp. 81-83). Socrates's insistence upon painstaking analysis signaled for Nietzsche not only the end of the vitality of Greek culture, but also the beginning of an age of men with diminished spirits dependent upon rational analysis rather than creative myth.

An even more blistering attack, though from a different angle, comes from the late journalist I. F. Stone (1988). Stone sees Socrates as democracy's enemy, one who believed that the herd of men needed to be firmly ruled (p. 38). This political view, coupled with the belief that knowledge is absolute and unattainable, and that virtue and knowledge could not be taught (pp. 63f.), makes it difficult for Stone to see how

Socrates could be defended as a teacher or even citizen of Athens. Stone's book made a splash in the trade press because he attempted to defend Athenian democracy against Socrates.

A more measured critique of Socrates's influence can be found in Bruce Kimball's widely discussed book, Orators and Philosophers (1986). Kimball points out that the philosophical tradition of Socrates has won out in contemporary liberal education over the oratorical tradition of Cicero. Kimball sees a tension between the pursuit of knowledge on the one hand, and the recognition and maintenance of the importance of historical traditions within learning communities on the other hand.

Kimball's discussion has crucial educational import. It challenges us to find ways to keep alive the Socratic spirit, however corrosive or parasitic of myth and culture, while also maintaining an appreciation and a cultivation of tradition and custom advocated in the Ciceronian oratorical view. This challenge was of course Nietzsche's own too, made clear in The Use and Abuse of History. We shall keep this theme from Kimball and Nietzsche in mind as we examine contemporary Socratic pedagogy.

Such views were not voiced specifically apropos of educational practice; yet they are key to understanding Socratic pedagogy today. This import is evident in a heated published exchange between Richard Paul and Louis Goldman concerning the role of Socratic inquiry in the schools (Goldman,1984; Paul,1984). Goldman believes that Socratic questioning can be dangerous if begun too early: "A proper education of the young must begin with a firm grounding in the nature and values of our culture" (Goldman, 1984, p. 60; see also Nietzsche 1872, 1956; Beatty, 1984; Kimball, 1986). He notes that Plato advocated dialectics only after a long preparatory education. Socratic questioning can become dynamite in the wrong hands, and we only approximate his practice (p. 62). Goldman recommends that we attend to traditional (Ciceronian, in Kimball's term) education for the young, and not encourage too early an introduction to dialectics.

Richard Paul, perhaps the most well known advocate of critical thinking in the schools, disagrees with Goldman. He believes that we must foster the habit of thinking critically at the same time and in tandem with an appreciation of culture. He takes up the challenge offered by Kimball and others. To borrow Kimball's terms, Paul believes that a synthesis of Socratic inquiry and Ciceronian traditionalism should be fostered. Paul goes further by making a claim common to theorists interested in philosophy for the young (Matthews, 1980; Lipman, Sharp, & Oscanyan, 1980). Thinking philosophically occurs naturally in children. Infectious curiosity manifested in childlike wonder and the persistent questioning that attends such wonder should be harnessed by a sensitive teacher to further the appreciation of cultural traditions and other educational aims.

Socrates as Teacher: A Reexamination

Prominent philosophers of education extend these perspectives, from Nietzsche to current debates in the area of critical thinking. Perhaps the most sustained attempt to grapple with the legacy of Socrates for pedagogy has been made by Sophie Haroutunian-Gordon and her colleagues (Hansen 1988; Haroutunian-Gordon 1987, 1988, 1989, 1990, 1991; Haroutunian and Jackson 1986).

Through a close reading of several of the Socratic dialogues, particularly the Gorgias, Meno, Philebus, Phaedo, and Protagoras, Haroutunian-Gordon challenges the notion of a "Socratic method." For instance, she points out Socratic inconsistencies that call into question use of the term. She makes a further claim that the reason Socrates does not follow a prescribed formal method is that he is in what educational researchers now call an "ill-structured teaching situation." Following a predetermined dialectical blueprint will not suffice for the way that a discussion may have gone "awry" (Haroutunian-Gordon, 1988, p. 231). In such situations, the questions one asks depend on the content of the conversation, and how nuance and shadings of meaning issue forth their own structure. Certainly many post-Wittgensteinian philosophers, as diverse as Grice and Gadamer, have explored this phenomenon long known to writers of imaginative literature. For our purposes, so too have experienced teachers known how to work their way out of ill-structured situations to effect learning for diverse students.

Elsewhere, by way of showing again the inadequacy of a formal description of Socrates's teaching, Haroutunian-Gordon attempts to "identify pedagogical aims" (1987, pp. 119f.) by giving four suggestions about what these might be: 1) bring interlocutors to aporia; 2) pursue truth about fundamental questions; 3) teach proper intellectual habits; 4) modify the moral principles of the interlocutors. Though Socrates may advocate the philosophical life via these aims according to Haroutunian-Gordon, he does not demand that others follow this life, nor are these purported aims necessarily relevant to the "task of explaining why he did what he did in the dialogues" (1987, p. 129). Haroutunian-Gordon's arguments are important, for they powerfully undermine an easy mimicry of the use of Plato's Socrates in one's pedagogy.

Socratic Pedagogy in the Meno

The Socratic legacy offered up by Haroutunian-Gordon and her colleagues, along with the views discussed by recent commentators, make it difficult to see how Socrates has become such a pervasive pedagogical model. He says repeatedly that he is not a teacher, and then seems almost intent on proving that claim by irony, inconsistent action, and an occasional long-winded speech, as at the end of the Gorgias. Yet, perhaps we can turn to one place where many have looked when they speak of Socratic teaching, Plato's dialogue Meno. An old man drawing geometric figures in the sand for a young slave boy is a powerful image of what many believe Socratic teaching to be.

Nevertheless, we must be careful with this seemingly transparent instance of pedagogy. Though an important theme of the dialogue comes when Socrates extracts the distinction between knowledge and true opinion through coaxing and vivid imagery, his supposed drawing out of the recollected geometric wisdom from the slave boy is troublesome as a display of pedagogy.

Socrates begins his lesson by putting words in the mouth of the slave boy (82B f.). Is this a convincing display of pedagogy? Leaving aside the blatant (to my eyes at least) problems of power and dominance of an elderly Greek citizen teaching a slave boy, this example of teaching has always left me cold. It is not apparent at all that teaching has occurred though it is a convincing display of inference as R. E. Allen (1959) has pointed out. It is not made clear in the dialogue that the slave boy is somehow capable of using his knowledge. He appears more like a sounding board for Socrates, who here seems to be just a mouthpiece for the theories of recollection (anamnesis) and innate knowledge.

I grant that a more generous reading of the Meno that sets aside this power differential is possible (Macmillan and Garrison, 1988; Burbules, 1993, pp. 120-122). Here Socrates is actually teaching when he asks his leading questions of the slave boy, because these questions are disguised answers to the questions that the boy should be asking. These questions are "stepwise" (Burbules, 1993, p. 122) in an instructional kind of dialogue where the end is known by the teacher, and implicitly known by the slave boy (Macmillan and Garrison, 1988, p. 154).

Contemporary Inspiration: Capturing, and Missing, the Socratic Spirit

Though the Meno may be troublesome to some as pedagogy, it has provided inspiration to classroom teachers. The famous passage (80A-B) where Meno chides Socrates for being like the electric ray (or torpedo) for delivering perplexing questions has provided Donald Thomas (1985) with a way to teach so that students will go out on their own and dig under the surface. In a brief and thoughtful essay, Thomas describes an episode in his early secondary school teaching career when he dramatically presented a sermon by the Puritan preacher Jonathan Edwards for his students. He wanted to stun his students into a perplexity that might be uncomfortable, much as Socrates makes Meno uncomfortable with his persistent questions. Thomas wanted them to see Edwards come alive so that these contemporary students would not forget the Puritan's images. The "torpedo's touch" was there, much to the chagrin of a team of behaviorally oriented evaluators in the back of the room.

Many years after this incident, Thomas still uses the "torpedo's touch" in his pedagogical array. Like Socrates, he often begins with pleasantries and surface talk, waiting for the right moment to deliver the stark and perplexing questions that may provoke wonder coupled with a realization of ignorance in his students (p. 222). Yet, Thomas's essay is too brief for him to give us examples of his questions, and to recreate a number of different pedagogical scenarios. Furthermore, we would want to know just how his questions were akin to those of Socrates beyond being perplexing and intellectually numbing.

While Thomas has taken inspiration from Socrates in his classroom practice, others attempt to devise Socratic teaching strategies devoid of such spirit. Some of the most flagrant "abuses" associated with using Socrates as a pedagogic model come when superficial aspects of the Platonic Socrates are used uncritically as pedagogic strategies.

Fishman (1985) notes several of these "misconceptions." The Socratic method is often seen and used today as an open-ended question and answer process (p. 185). Kay and Young (1986) equate Socratic teaching with asking more questions in the classroom and with the encouragement of students to become independent and autonomous thinkers. They compare Socratic questioning with a teaching strategy called "ReQuest," developed by the educationist Anthony Manzo. No mention of content or aim of the questions is given by Kay and Young; apparently to them it seems sufficient that the teacher is a full-time questioner in order to be dubbed Socratic.

Beyond Inspiration: Current Socratic Teaching

In what follows, I explore a number of examples of Socratic teaching strategies that have gone beyond either drawing inspiration from the dialogues or missing that inspiration. A weakness in Thomas's approach is that a pedagogical strategy, rich with examples, is not spelled out in his brief essay. On the other hand, if an understanding of the Socratic mission is absent, we may be led to the lifting and distorting of formal qualities of Socratic practice in our teaching, as with Kay and Young (1986). The teaching of Vivian Gussin Paley and another Chicago based Socratic practitioner, Mortimer Adler, provide our next examples as we enrich our horizon concerning Socratic pedagogy.

Vivian Paley, recently retired from years of teaching kindergarten, tells us of numbness of a different kind from that of Thomas's "torpedo's touch." She describes candidly her lack of interest and enjoyment in her early years of teaching (1986). She happened to observe a colleague using the "old Socratic method" (p. 123) she too had once used as a Great Books discussion leader. Then she began to realize how excited she was about the process of thinking going on in the minds of her students. She now affirms the place of this process over any other outcome, or product, in her teaching (1990). Children are not interested in answers, she claims, but are fascinated by process (1990; see also Matthews, 1980) and entranced by meaning making and the play of language.

The impetus for renewed interest and curiosity about her own teaching came from the hard realization that she did not know the answers to the questions that her young charges were posing. Paley was thus forced to keep asking relevant questions, based not on her own preconceptions, but rather on how the child was thinking about a topic (1986, p. 124). The classroom drama, in which her students enacted imaginative stories of their own construction, became for her "a paper chain of magical imaginings mixed with some solid facts" (1986, p. 123). This paper chain offered Paley abundant opportunities for her version of Socratic probing.

Yet Paley the teacher goes beyond Socratic questioning in the classroom. She turns the questioning reflexively upon herself and her own thinking with a "specific tool" (1990) she has used for years, the tape recorder. Paley tapes daily ninety minutes of her students' stories and the accompanying dialogue. The tape recorder, with its "unrelenting fidelity" (1986, p. 123) trained her to listen precisely to what the children say. In transcribing the taped dialogue, large chunks of which appear in her books, Paley has the opportunity to review all that went on in the classroom. Using what she calls an "internalized Socratic method" (Obermiller, 1986, p. 19), she takes herself to task in preparation for her writing, asking herself questions like "why did I ignore that question?" or "is that something I could have taken up with him?" (1990).

For Paley, this activity is part of the "intellectual game of teaching" (1990). Interacting with preschoolers as they play with blocks is not merely play, but also the process of thinking and intellectual inquiry. Her exhausting teaching, taping, and transcribing regimen is an important living manifestation of the Socratic notion of the worth of the examined life. This element of reflexive inquiry aimed at self-knowledge, difficult to achieve, is absent from such purported Socratic teaching advocated by practitioners like Kay and Young.

Paley's methods have attracted attention and acclaim. Yet Mortimer Adler and his supporters espouse an even more widespread version of Socratic teaching. Adler's Paideia Proposal (1982) is one of the key documents of the recent school reform movement. In this brief work he advocates three interrelated ways of learning (p. 23) that should be followed by all students regardless of age or ability: 1) the acquisition of knowledge by lectures, memorization, and other means; 2) the development of intellectual skills, through coaching; 3) the enlargement of understanding through Socratic discussion of ideas and texts. However, the overwhelming majority of the focus given in the implementation of the Paideia Proposal, has been on the third type of learning, the Socratic seminar (Gray, 1988, Sizer, 1984, Chapter 5). Let us turn to a discussion of Adler's version of Socratic pedagogy.

Adler's description of seminar pedagogy is deceptively simple: A "discussion in which students both ask and answer questions" (1982, p. 53). One of his close associates, Patricia Weiss, defines a seminar as: "(an) educationally oriented discussion in which ideas, issues, or principles are examined...The main teaching method used in seminars is one of questioning and examining responses. This style of teaching is often referred to as Socratic teaching, named after Socrates who used questions in his teaching of the youth of Athens in 400 BC" (1987, p. 1; emphasis added).

Weiss then describes the three tasks of the seminar leader proposed by Adler: " 1) to ask a series of questions, 2) to examine the answers by trying to draw out the reasons for them, or their implications, 3) to engage the participants in a two-way talk with one another when views appear to be in conflict" (p. 1). I can recognize Socrates in 1 and 2, though I cannot recall anywhere in the dialogues where Socrates encourages his interlocutors to debate each other. Rather, with few exceptions (Callicles comes to mind), these interlocutors are more likely to give monosyllabic replies to Socrates's withering questions, prompting more than one reader to wonder just how dialogic these accounts were intended to be.

One the other hand, though this Adlerian technique may not be true to the Platonic Socrates, might it be seen as a commendable development of Socratic practice? After all it does seem odd (unless you consider Plato's own agenda for his created characters) that these interlocutors, many of whom are absurdly laconic, rarely argue amongst themselves (the Gorgias being a striking exception). Sadly, though, at least in my repeated observation of seminars led by Adler himself and some of his associates, this third task of a seminar leader is as rarely practiced today as it might have been in ancient Athens.

Let us now turn to a closer examination of how Weiss practices

Socratic teaching. In her manual that accompanies the videotapes of Adler leading seminars for high school students (1987), Weiss provides a detailed discussion of how to structure a seminar. She suggests that the teacher first set an atmosphere that will allow students to feel at ease in asking questions. This may include putting to one side any expertises students may bring to the text at hand (Gray, 1988) so that general discussion among (near) equals may be established. Weiss begins her classes with a variety of nonthreatening questioning techniques (e.g., round robin, voting, random call on whether students like or dislike Socrates are typical in her teaching of the Apology) to get discussion going.

Once discussion is underway, Weiss may move to ask whether Socrates is a teacher, opinions on the charges made against him in the Apology, or whether he is guilty or innocent. Like Haroutunian-Gordon (1988), Weiss acknowledges the "ill-structured teaching situation" through this emphasis upon making teachers aware of the importance of being prepared to ask unscripted follow-up questions (Weiss, 1987, p. 2). These practices are all commendable, but they rest upon a crucial assumption, made clear by another Paideia advocate, Dennis Gray. Even as Gray asserts that Socrates had no syllabus, he declares that the purpose of Socratic teaching is to focus always on texts, with even the opening question based upon a close study of the text at hand (1988).

The changes we moderns have made in the name of Socrates could not be clearer. Socrates, of course, did not use a common reading around a seminar table. Furthermore, this assertion by Gray makes apparent another related assumption of the Paideia method, namely that great works will contain great ideas.

The Dark Side of the Socratic Legend

The image of Socratic teaching presented above has been mixed. Socrates can be difficult and disarming. Yet we educators are often intent upon seeing Socrates in the warm glow of history as the one who began humanistic inquiry. In this section, I shall return to an unromantic view of Socrates that I have so far only presented through other writers such as Stone and Nietzsche.

I shall suggest the importance, though with qualifications, of the dark side of the Socratic spirit by turning to some first-hand accounts of legal pedagogy, and the use of the "Socratic method" in law schools. In spite of Adler's inroads into the nation's schools, the popular image of Socratic teaching often comes from the so-called "Socratic method" used in law schools. I gleaned insight from colleagues from graduate school who hold the doctoral degree in philosophy and have also studied law.

Many of us have never entered a law class, but we feel that we know what goes on there. We have seen John Houseman's portrayal of Professor Kingsfield in the film and television show, "The Paper Chase." Houseman's depiction of an unforgiving taskmaster asking his often-timid students withering questions is the beginning and the end of legal pedagogy for most of us, and for our perceptions on how Socrates is used in legal teaching. In consulting two colleagues who have experienced legal pedagogy, I was able to deepen my understanding of Socratic legal teaching beyond this popular image.

Peter Suber, professor of philosophy at Earlham College, holds both the PhD and JD degrees from Northwestern University. His description of a law class is truly harrowing: "Incorrect answers, undue delays in answering, or overt signs of nervousness are punished with sardonic jibes or withering glances. The atmosphere is humiliation; the punishment is humiliation...The consensus among students is that the method is not 'educational' in any traditional sense. It does not help one learn cases or legal reasoning. It is sadistic" (1990). Suber sees ample evidence in the dialogues to think that Socrates behaved similarly. Furthermore, Suber believes that the so-called legal Socratic method is used in different ways in law schools of different levels of prestige (1990). In the most prestigious category, students behave in the "Paper Chase" fashion, reciting the facts and attendant arguments while standing and attempting to answer the professor's questions.

On the other hand, Suber notes what he calls second echelon schools and below may be places where the method is more humane. Here there may be more emphasis upon reasoning and thinking rather than performance. Unlike the first instance cited, this gentler use of the method may in fact emphasize "respond(ing) to well-crafted counterfactuals again and again" (1990) in an atmosphere of support and trust.

Another former colleague, Mark Olson, also holds the doctoral degree in philosophy from Northwestern University and a law degree from the University of California at Berkeley. Olson, now a practicing attorney, takes a different tack in discussing his experience. He begins by offering a definition of what he carefully calls the "legal Socratic method: "(It) employs the use of actual recorded court cases to teach students the rules of law and their application and justification (whether clear or not, whether persuasive or not), through the instructor's use of a series of hypotheticals based on the main case and through the students' discussion of the case and the hypotheticals. Its successful use and reception calls for skill and wit" (1990).

Olson reminds us of other factors that I agree are crucial to the understanding of the legacy of Socrates for pedagogy. The Socratic method evolved in law training as a "historical formation, which, in its present form presupposed the existence of a legal casebook" (1990). Above all for Olson, the method is not a technique; when it is so practiced it is characteristic of inept instructors. In those classes students are not probed, but are allowed to give "unreflective (knee-jerk) responses to complex social issues" (1990).

One of my deep seated and cherished beliefs has again been questioned by this knowledge. I want to believe, along with Adler and other sanguine educators, that Socratic teaching is a means to search for truth. I still muse in uncritical moments about a Socrates, beneficent and maligned, leading the youth of Athens on the golden path of instruction. Even in graduate programs in philosophy one does not discuss often the darker side of this practice as argued by Stone and Nietzsche, and brought to the fore here in a different way by Suber and Olson. Stone's criticism of Socrates is too recent; besides, he built his reputation as the consummate outsider journalist who only taught himself Greek in his waning years. He did not belong to the anointed academic club of classical scholarship. Nietzsche, though a classical scholar, is often dismissed as a German at best and a raving crank at worst, particularly when it comes to his views on Socrates.

Yet, this "darker" side of Socrates must be preserved, as I shall contend in the following section, if we are to truly "use" and not "abuse" Socrates in present day teaching. Suber's description of a harrowing law class may be an extreme version of such practice, though a recent feminist critique of legal education supports Suber's claims. Guinier, Fine, and Balin (1997) go even further to call the legal Socratic method "ritualized combat" that is harmful and counterproductive to the education and well-being of women law students.

Suber's description of the "sadistic" querying that may go on in higher echelon law schools may be true to Socrates in one sense; he was relentless and oftentimes unpleasant. But we must ask to what end these displays are headed. In the following section, I shall seek to show that we must preserve the wily, irascible Socrates most of us have come to love (or hate) so that at the same time we preserve the core of his mission.

Conclusion: Determining the Use and Abuse of Socrates

We have seen how Socrates is part of many classroom situations, from Paley's kindergarten on up to law school. Which of these are legitimate uses of Socrates and which are abusive? To determine such appraisals, I believe we must use several standards. Abuse of Socrates does not necessarily come, as might be first thought, when the Socratic "victim" is mischievously questioned and pierced with sardonic barbs. Abuse may come rather more from well-meaning educators who, perhaps in the joy of discovering a technique that is liberating and aims toward thinking, strip Socrates of his power. How could the Socratic legacy be so diminished?

First, we may forget that Socrates at his best was attempting to uncover self-knowledge and encouraging others to do so too.. He followed his daemon and eschewed followers. As both Stone and Suber underscore, Socrates was devious and crafty. These factors must lie at the core of any interpretation of Socrates for present day teaching. If we apply (and I use this term deliberately) a Socratic method to any topic, this strategy does not necessarily guarantee that self-knowledge will occur. Self-knowledge is a difficult concept, as the irony used by a Socrates or a Soren Kierkegaard seems to suggest. Yet to forget the unsavory aspects of Socrates is to forsake the Socratic spirit, and thus to abuse the legacy of Socrates for education.

A related abuse of Socrates in present day teaching comes when we believe uncritically that Socrates himself was a teacher. The word teacher makes most of us who are in the "education business" think of someone who may devise and implement a curricular rationale. If Socrates was indeed a teacher, then he must have had a specific pedagogy and a specific set of topics that can be learned by others, the reasoning goes. A central insight of Burbules's recent book on dialogical teaching (1993) is in seeing clearly Socratic teaching as a multifarious repertoire. Burbules argues that Socratically inspired teachers play a dialogical "game" that, though guided by rules, is sensitive to context. Plato's own writing of the dialogue is itself a Socratic teaching act. The writer is the teacher, and the reader, the student, both of the dialogue, and, reflexively, of him or herself. But this has not deterred other educators from advocating what they suppose are simple teachable strategies and curricular objectives derived from Plato's character.

While I have observed Mortimer Adler using irony and even humiliation in a manner akin to Socrates, it is not clear that those trained by him have the confidence or the temperament to use these ploys. I have watched well-intentioned teachers trained under Adler leading supposedly "Socratic" discussions without suggesting even a hint of irony or challenge, something Adler himself criticizes as "watered-down" seminars (Adler, 1990). Perhaps a good number of teachers find themselves incapable of being "mischievous, disingenuous, and cunning, and occasionally even devious" (Suber, 1990) in the way that Plato's Socrates was.

Conflicts between Socratic teaching and other aims of education are also apparent and disturbing. Educators are urged to be supportive, to nurture their students, many of whom are currently "at-risk." Teachers must often serve as surrogate parents to students from dysfunctional families. I have taught Plato's Socrates to a number of graduate students in special education. They have all told me that using Socratic dialogue moves in their teaching would be highly problematic if not impossible. It is thus difficult and perhaps even at cross-purposes to use such a pedagogical method and encourage the cultivation of self-knowledge with these students.

Here is where the sensitivity, knowledge, and skill of a teacher, well-versed in a Socratically inspired repertoire of pedagogical strategies and moves in the dialogue game (Burbules, 1993), comes into play. In addition, such a teacher must have a sympathetic understanding of each student and the nuances of that particular classroom climate. Otherwise such abuses already discussed such as either a stripping of Socratic teaching to just a questioning exercise, on the one hand, or the "ritualized combat" experienced by many women in law classes, could occur.

My argument points to the need to recognize an enduring core of the Socratic legacy for teaching. Haroutunian-Gordon and others have given enough textual evidence in order for us to be suspicious of thinking that Socrates was a teacher in any conventional or current sense of that term. Other commentators as diverse as Nietzsche, I.F. Stone, Bruce Kimball, and Louis Goldman have called attention to the corrosive and even dangerous qualities of Socratic inquiry. Yet why does Socrates continue to leave the torpedo's deep marks upon most anyone who reads the dialogues, and on those of us who are inspired to model his actions in our own teaching?

The Socrates of Plato's dialogues continually cuts past areas of knowledge apprehended by either episteme or phronesis, theory or practice. Socrates can make us feel that the failure to sustain a thesis or find a definition is not just a defeat of intelligence, but rather a moral disaster (Vlastos, 1971, 1980, p. 6). Socrates may not have given us a simple "method" that we can apply to any topic, and it may be difficult to teach Socratically in today's "antidialogical" schools (Burbules, 1993). Yet the larger issues raised in the dialogues must not be ignored. If anything, Socratic irony confounds many of the simplistic interpretations of the Socratic legacy for teachiing. The care of the soul, the project of moral inquiry, and a searching that cuts across social class should be the first and foremost use, and ultimate worth, of Socrates for present-day teaching (Vlastos 1971, 1980; see also Gadamer, 1986; Seeskin, 1987; Johnson, 1989).

Note: This paper is based on a presentation at AERA. I wish to thank Jim Garrison, Sophie Haroutunian-Gordon, Michael Jones, and Ralph Page for comments upon several subsequent drafts prior to submission, and Gene Glass and four anonymous reviewers for this journal for their comments subsequent to submission.

References

Adler, M. J. (1982). *The paideia proposal.* New York: Macmillan.

Adler, M. J. (1990, January/February). No watered-down seminars. *The Paideia Bulletin*, VI(3), 1; 6.

Allen, R. E. (1959). Anamnesis in Plato's Meno and Phaedo. *Review of Metaphysics,* XIII(1), 165-74.

Beatty, J. (1984, winter). The complexities of moral education in a liberal, pluralistic society: The cases of Socrates, Mrs. Pettit, and Adolf Eichmann. *Soundings*, 67(4), 420-42.

Burbules, N. C. (1993). *Dialogue in teaching: Theory and practice.* New York: Teachers College Press.

Dannhauser, W. J. (1974). *Nietzsche's view of Socrates.* Ithaca, NY and London: Cornell University Press.

Fishman, E. M. (1985, fall). Counteracting misconceptions about the Socratic method. *College Teaching*, 33(4), 185-88.

Gadamer, H-G. (1986). *The idea of the good in Platonic-Aristotelian philosophy* (P. C. Smith, Trans.). New Haven, CT: Yale University Press.

Goldman, L. (1984, September). Warning: The Socratic method can be dangerous. *Educational Leadership*, 42(1), 57-62.

Gray, D. (1988, summer). Socratic seminars: Basic education and reformation. Basic Education: Issues, Answers, and Facts, 3, 14. Washington, DC: Council for Basic Education.

Guinier, L., Fine, M., & Balin, J. (1997). *Becoming gentlemen: Women, law school, and institutional change*. Boston: Beacon Press.

Hansen, D. T. (1988, spring). Was Socrates a "Socratic" teacher? *Educational Theory*, 38(2), 213-224.

Haroutunian-Gordon, S. (1987, fall). Evaluating teachers: The case of Socrates. *Teachers College Record*, 89(1), 117-32.

Haroutunian-Gordon, S. (1988, spring). Teaching in an "ill-structured" situation: The case of Socrates. *Educational Theory*, 38(2), 225-237.

Haroutunian-Gordon, S. (1989). Socrates as teacher. In P. W. Jackson & S. Haroutunian-Gordon (Eds.) *From Socrates to software: The teacher as text and the text as teacher* (Eighty Ninth Yearbook of the National Society for the Study of Education, Part I, pp. 5-23). Chicago: University of Chicago Press.

Haroutunian-Gordon, S. (1990). Statements of method and teaching: The case of Socrates. *Studies in Philosophy and Education*, 10, 139-156.

Haroutunian-Gordon, S. (1991). *Turning the soul: Teaching through conversation in the high school.* Chicago: University of Chicago Press.

Haroutunian, S., & Jackson, P. W. (1986, fall). The teacher in question: A study of teaching gone awry. *Teachers College Record*, 88(1), 53-63.

Johnson, T. W. (1989, fall). Teaching as translation: The philosophical dimension. Thinking: *The Journal of Philosophy for Children*, 8(3), 34-38.

Kay, L. H., & Young J. L. (1986, July-August). Socratic teaching in social studies. *Social Studies*, 77(4), 158-61.

Kimball, B. A. (1986). *Orators and philosophers: A history of the idea of liberal education.* New York: Teachers College Press.

Lipman, M., Sharp, A. M., & Oscanyan, F. S. (1980). *Philosophy in the classroom.* Philadelphia: Temple University Press.

Macmillan, C. J. B., & Garrison, J. W. (1988). *A logical theory of teaching: Erotetics and intentionality.* Dordrecht, the Netherlands: Kluwer.

Matthews, G. B. (1980). *Philosophy and the young child.* Cambridge, MA: Harvard University Press.

Nietzsche, F. N. (1872, 1956). *The birth of tragedy* (F. Golffing, Trans.). Garden City, NY: Doubleday.

Nietzsche, F. N. (1874, 1979). *The use and abuse of history* (A. Collins, Trans.). Indianapolis, IN: Bobbs-Merrill.

Obermiller, T. (1989, summer). All in a day's play. *University of Chicago Magazine*, 15-19.

Olson, M. A. (1990, January). Personal correspondence with the author.

Paley, V. G. (1986, May). On listening to what the children say. *Harvard Educational Review*, 56(2), 122-131.

Paley, V. G. (1990, January). Telephone interview with the author.

Paul, R. W. (1984, September). The Socratic spirit: An answer to Louis Goldman. *Educational Leadership*, 42(1), 63-64.

Pekarsky, D. (1994). Socratic teaching. A critical assessment. *Journal of Moral Education*, 23(2), 119-134.

Plato (4rth Century BC, 1973). *The collected dialogues of Plato* (E. Hamilton and H. Cairns, Eds.). Princeton, NJ: Princeton University Press.

Seeskin, K. (1987). *Dialogue and discovery: A study of Socratic method.* Albany, NY: SUNY Press.

Sizer, T. R. (1984). *Horace's compromise: The dilemma of the American high school.* Boston: Houghton Mifflin.

Stone, I. F. (1988). *The trial of Socrates.* Boston: Little, Brown.

Suber, P. (1990, January). Personal correspondence with the author.

Thomas, D. W. (1985, May). The torpedo's touch. *Harvard Educational Review*, 55(2), 220-22.

Vlastos, G. (1971, 1980). *The paradox of Socrates*, in G. Vlastos (ed.) *The Philosophy of Socrates: A Collection of Critical Essays*, (pp. 1-21). Notre Dame, IN: University of Notre Dame Press.

Weiss, P. F. (1987). *Great ideas: A seminar approach to teaching and learning.* Chicago: Encyclopedia Britannica.

About the Author

Anthony G. Rud Jr.

rud@purdue.edu

Associate Dean
School of Education, 1440 LAEB
Purdue University
West Lafayette, Indiana 47907-1440 USA

email: rud@purdue.edu
Phone: 317-494-6542

Fax: 317-494-5832

Home page: http://www.soe.purdue.edu/faculty/rud.html

Anthony Gordon Rud Jr. is associate dean, and associate professor of educational studies, in the School of Education at Purdue University. He received his A.B. with honors from Dartmouth College, and his M.A. and Ph.D. in philosophy from Northwestern University. Rud is a founding member and former officer of the Association for Philosophy of Education, a former member of the Committee on International Relations of the American Educational Research Association and the Committee on Pre-College Instruction of the American Philosophical Association, and a senior associate of the Council for Basic Education. He has served as a consultant for school systems on leadership issues, critical thinking, moral education, and school reform, and for organizations as diverse as the National Paideia Center at The University of North Carolina at Chapel Hill, the Center for the Advancement of Ethics and Character at Boston University, and the Department of Special Education at the University of South Florida.

Rud came to Purdue from The North Carolina Center for the Advancement of Teaching, where he served as Senior Fellow. He also has worked at Dartmouth College as an administrator, adjunct faculty member, and freshman advisor, and taught high school humanities and social studies. Rud's research interests center upon the intersection of the foundations of education and educational practice, with particular emphasis upon the preparation and professional development of teachers and educational leaders.

Rud is senior editor and contributor to *A Place for Teacher Renewal: Challenging the Intellect, Creating Educational Reform*, published by Teachers College Press. The author of a number of articles and reviews, he regularly makes presentations at major professional conferences. Rud serves on the editorial boards of several academic and professional journals, and is the chair of the editorial board of Purdue University Press. He joins James W. Garrison as coeditor and contributor to a volume of essays entitled *The Educational Conversation: Closing the Gap*, published by the State University of New York Press in 1995.

Rud lives in West Lafayette, Indiana with his wife Rita and daughter Rachel.

Copyright 1997 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://olam.ed.asu.edu/epaa

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey@olam.ed.asu.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Marshall University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Rocky Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
Arizona State University

Dennis Sayers
University of California at Davis

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina--Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher Education

Mary P. McKeown
Arizona Board of Regents

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois--UC

Robert T. Stout
Arizona State University

---

archives | abstracts | editors | board | submit | comment | subscribe | search

EPAA Commentary

Contributed Commentary on Volume 5 Number 20: Rud *The Use and Abuse of Socrates in Present Day Teaching*

---

30 May 1998

Rick Garlikov

Rick@Garlikov.com

Preface

I am deeply troubled by the comments about Socrates and the Socratic Method that Anthony G. Rud, Jr. makes in his "The Use and Abuse of Socrates in Present Day Teaching" (EPAA Volume 5 Number 20, November 24, 1997), because I think they reflect what many educators today might feel about Socrates as portrayed in the Platonic dialogues; and if so, I think that is a terrible loss for today's students. In this response I want to deal only with issues about the Socratic teaching method, not with the interesting, and I think false, claim Rud seems to accept (ascribing it originally to Nietzsche) about the supposed conflicts between "rationality" and "creative myth" or between the search for truth and the continuing of cultural traditions. And I do not wish to consider specific beliefs of Socrates involving possible errors in reasoning or in the factual knowledge from which he was reasoning. I wish only to examine issues relevant to instructional methodology.

What I find particularly troubling is that Rud (and, from his citations, apparently many others) doesn't see what seems to me to be quite obvious in regard to what Socrates' methodology is all about.(Note 1) The method of questioning of the slave boy is not unlike Socrates questioning of *Euthyphro* or of the many other, often famous and powerful, adults in the dialogues. So where Rud sees it as displaying "blatant...problems of power and dominance of an elderly Greek citizen teaching a slave boy," I see the age and class difference as irrelevant features of that dialogue, since the essential features in it are no different than in the others, where Socrates questions quite prominent and highly positioned citizens. Nor, in many cases are the questions, and the answers given by the characters in the dialogues, very different from what would occur today if one were talking to someone about those very same subjects in a modern vernacular and with references to contemporary examples of the same sorts of actions and ideas Socrates gives his examples to illustrate.

In discussing both Burbules and Haroutunian-Gordon, Rud seems to think it somehow surprising that Socrates does not follow prescribed formal methods and that he did not have a "specific pedagogy and a specific set of topics that can be learned by others." I do not know how to account for finding that surprising. It is quite clear from the dialogues that

Socrates simply had thought deeply about a great many different things, and that he questioned people who expressed views he thought false about those things to see whether it was he or they who were in error, probably thinking it more likely they would be the ones who were mistaken, but entertaining the possibility he was the one who was mistaken. In the dialogues I have read, it is quite clear that Socrates' analysis is always deeper and more sustained than the other participants, and that the latter tend to express either unanalyzed, unreflective, popular views or views for which they may have some shallow reasons, but not very good ones - certainly not ones that will withstand scrutiny. To characterize this sort of inquiry, where one responds to others' views by showing them the problems with those views, as an "ill-structured teaching situation" that teachers need "to work their way out of" is to think of teaching much too narrowly as something that can only occur with a particular topic in a particular formal and prescribed way under particular conditions regardless of what beliefs or knowledge or ability one's students bring to the classroom. And the requirement of a specific pedagogy and set of topics ignores the common notion of teaching as something, say, parents might do when they teach their kids table manners, bicycle riding, the rules of baseball, how to throw a curve ball, badminton, croquet, how to drive, checkers, card tricks, how to deal with defeat or disappointment, and a zillion other things that parents teach children without curricular rationales or systematic, formalized topics offered at specially chosen preconceived times.

The method involves teaching things at times they are relevant in ways that are most relevant to the person one is teaching, or guiding. That necessarily involves being able to ascertain what they already know or believe and understand that can help you teach them what you want them to learn. Although there will be similar elements each time you teach the same topic to different people, some quite different questions and ideas may be required as well, sometimes extremely different questions. One class I taught in the early 70's was so wedded to the principle of being true to yourself and honest with everyone else that when we discussed "When is it right to break a date, and why?" they honestly believed one should break a date, or do **any**thing whenever one feels like it, regardless of what expense or effort the other person might have put into preparation for the date, and regardless of how important it might be to him or her. I thought of every possible question I could to get them to see that was not really a very good principle and that they really couldn't possibly believe it, but no matter what case I presented to them, they consistently took what seemed to me to be the most absurd and untenable position in favor of "doing whatever you really want to" even if it meant committing murder you thought you could get away with. So I had to use a very different strategy to get them to see they didn't really believe that principle was a good one to follow. That was the only group of students I ever taught who held that view; and so the way I had to deal with them during those classes was different from the way I ever had to deal with any other group. I don't consider such classes ill-structured teaching situations. I consider them to be normal teaching situations if one is really trying to address students' current knowledge, understanding, beliefs, and sincere interests as a starting place to help them learn.

And that makes reading Socratic dialogues harder than participating in one. When the interlocutor gives an answer you would not have given, immediately the dialogue will veer from the direction it would have taken had you been the one answering. You may lose interest or you may not understand or appreciate the answer; or you may not "get" the teacher's response to that answer. Yet you are not there to ask him what s/he means or why the question is relevant. Further, when you are not a participant, it may be difficult to see certain things that are important to see in order even to understand the questions. In the edition I have of the Meno, there is a diagram of the squares Socrates is discussing with the slave, but it is a finished diagram, containing all the elements Socrates draws out for the slave boy one at a time as he needs them. It is difficult to tell which lines Socrates is pointing to as he asks questions about the various squares and their areas relative to each other. So what might be quite instructive for the slave boy in Socrates' presence as he draws or points to certain lines in the figure might be more difficult for a reader to understand than would be reading a straightforward essay or argument. Unless the answers given by another student are the same as what you would have given, the reading of a Socratic dialogue involving him or her is **not** the same educational experience as participating in such a dialogue. Even for the same topic, each occurrence of Socratic teaching is potentially quite different from another. That may explain why someone might see given dialogues as just a bunch of persistent questions that are "devious" and difficult. They may not appreciate the point or rationale for the specific questions in a specific dialogue. But since Rud is a philosopher, presumably knowledgeable about some of the topics in the dialogues, that does not explain to me **his** characterization of them that way. I would have thought he would understand why Socrates chooses the particular questions he does at the times he does.

Before getting into the method itself, I would like to dispose of two psychological characteristics Rud associates with it in some cases, that I think are also irrelevant to it: 1) ridicule of the respondent or of his answers, and 2) paralysis from fear by the respondent about having to participate in the enterprise. Ridicule can be used with almost any teaching method, even responding to raised hands, one can call on a student in a hostile and derisive manner. There is nothing about the Socratic method itself that requires hostility and sarcasm. And Socrates did not use it that way generally, if at all. To say or imply that because some people use the method "sadistically"to "humiliate" students in a form of "ritualized combat," the method has a "dark" or "unsavory" side ("to forget the unsavory aspects of Socrates is to forsake the Socratic spirit...", among other such passages) is like saying that because some people lecture in a boring manner, that lectures are by nature boring, or that because some jokes are tasteless that humor is by nature tasteless.

Further, neither the slave boy nor many of the other people Socrates questioned seemed to be numbed by fear of making some sort of mistake. Used in an engaging and kind way, there is no need for the method to be threatening. And while many students or colleagues tend to become suspicious "something is up" when someone, particularly a philosopher, launches into what seem to be Socratic types of questions, that seldom ties many tongues, unless one feels interrogated and threatened by the questioner for reasons other than his/her asking questions. Many of the characters in the dialogues do seem to find excuses to leave the discussions when they find their initial views untenable, but they don't seem hesitant to be in the discussions up to that point. For students, who are often not so wedded to a view that they find a successful challenge to it unbearable, the Socratic method is often quite exciting. My experience in using it with "at risk" or students with low GPA's and/or low self-esteem is that they find it interesting, challenging, stimulating, helpful, and nurturing. Rural Alabama students and urban inner-city students alike have said they wished their previous courses had been as thought-provoking, challenging, and attentive to their own ideas. Students often prefer to have their ideas taken seriously and disagreed with and questioned than to have them ignored or patronizingly given a good grade with no real critical or analytical attention paid to them. It is a terrible mistake to see the Socratic Method as being always and automatically antithetical to nurturing or supportive teaching. It only appears that way on the surface, or perhaps to people who don't think any ideas should be probed too deeply or their advocates challenged to justify them.

Finally in this preface, Socrates answers Rud's objection that "It is not made clear in the dialogue that the slave boy is somehow capable of using his knowledge." Socrates says what I think is true: "At present these notions have just been stirred up in him, as if in a dream; but if he were frequently asked the same questions in different forms, he would know as well as anyone at last." The way I tell this to people whom I teach photography is to say, "In two lessons [about four hours total time] I will teach you almost everything there is to know about photography so that you will understand it all. But you will not be able to apply it automatically or even keep it all in mind, for it will not have "sunk in". For that you will need to shoot pictures and bring them to me so that I can go over them and analyze them with you in light of the principles I will have taught you. That is when your understanding will start to sink in and become automatic(ally applied)."

The Socratic Method

The Socratic Method is simply a way, through asking certain sorts of pointed and stimulating, interesting or "live" questions, to get students, or anyone, to focus on the elements one thinks important for understanding of an idea or phenomena, in an order or perspective likely to help them gain that understanding. The idea is that by asking leading questions -- often logically leading questions-- about the elements one is bringing attention to, the other person will be able to see what you see about the idea or phenomena and will then attain the same understanding of it that you do.

The method only works with regard to logical/conceptual types of material; one does not learn "facts" via the Socratic Method, not directly anyway. (It may help the discovery of new facts by showing where one might need to "look" for phenomena, in the same way that theoretical physics often fosters discoveries in applied physics.) However, it can be used to make learning facts easier, by organizing their introduction to the student in a way likely to be more meaningful to him or her. More about this below, with regard to teaching photography.

When Socrates says it is not teaching, there are a number of things he might mean: (1) that it is not about telling anyone facts, or teaching them a skill (2) that it is not about telling anyone any thing s/he does not already know, in some sense of know, (3) that it only works with people who have some appreciation for, and sense of, inference, (4) that it is not about learning new things, but about putting into perspective things one knows but does not fully attend to or see the significance of, and/or (5) that it only applies to logical/conceptual aspects of material, not to the transfer of factual information . In the *Apology*, Socrates contrasts his kind of knowledge or wisdom with that of farmers, poets, etc. He does not have the kind of knowledge they have, which can be taught. He has a different sort of knowledge or insight. And he cannot "teach" one to have that, but he can demonstrate the results of his insight to others in certain ways, if they are willing and able to be receptive in certain ways. I consider what Socrates does to be teaching, even if he did not, because I consider methodically and intentionally fostering particular new perspectives and greater understanding in another person to be one (extremely important) form or aspect of teaching. But it is definitely not the same thing as trying to tell or teach facts or fact-based skills.

But the Method differs from just asking questions, especially open-ended or non-leading questions, or ones that one does not know the answer to oneself. That is in part why it is a teaching method, not just a brain-storming activity nor an activity whose point is merely to inspire thought or research by simply asking general or open-ended questions or questions one does not know the answer to oneself. Bill Hunter and I had a long e-mail debate, which is available in an edited form at http://www.Garlikov.com/teaching/dialogue.html about whether fostering research or discovery by students, via questions the questioner may not know the answer to, is teaching.

I use the Socratic Method for teaching many things. For a transcript of a fairly "pure" (i.e., questions only) use of the method, with commentary about it, for teaching a logical/conceptual idea see my "The Socratic Method: Teaching by Asking Instead of by Telling", http://www.Garlikov.com/Soc_Meth.html, But I also use the method for organizing material when I teach, say, photography; and for getting students to see mistakes in their reasoning in any subject matter area. An example of the latter case was in a discussion of homosexuality in an "Ethics and Society" course where many students said that homosexuality was wrong because (the idea of) it was so disgusting. I asked them whether they thought that such disgust was a sufficient characteristic to make an action be immoral. They said it was. I asked them then to close their eyes and think about ... their parents having sex with each other. They all let out an even bigger groan of disgust, and said they found that idea really disgusting. So I asked whether they would have to conclude then that it was immoral for their parents ever to have (or to have had) sex with each other. They agreed it was not. Of course they then asked whether that meant I thought homosexuality was moral. My response was that whether it is or is not is simply unrelated to whether it is personally disgusting or not to anyone. I was not trying to argue in this particular case for or against the morality of homosexuality, but was merely trying to get them to see that finding an action disgusting did not justify their thinking it must be immoral just because of that. Having introduced the point to them in that dramatic ("Torpedo's touch"?) way, we then go on to talk about other activities that they might sometimes characterize as disgusting but not immoral - such as dietary preferences that may be quite different from one's own, surgical procedures, etc.-- so that that one example is not seen to be either an anomaly or a complete explanation.

As to helping present facts in a manner conducive to seeing their significance, when I teach photography, I give certain demonstrations and then ask leading questions about them. Most people, even a great many who have taken a photography course or two, do not understand very well the significance of shutter speeds, aperture settings, or film ISO numbers. And they do not see that all these are related to each other in a simple way, nor what that portends for their picture-taking. So I open the back of an empty camera, remove the lens cap, and point the camera toward a light surface and trip the shutter, first at different shutter speeds, and then after some questions, at different aperture settings, letting them see what happens. After the shutter speed demonstration I ask them which shutter speeds let in more light, the ones where the shutter is open longer or shorter. They pretty clearly see that the longer the shutter is open, the more light comes in from a given source. After I point out that the shutter speed numbers are actually just reciprocals of time measured in seconds (e.g., 250 stands for 1/250 second, 2 stands for ½ second, etc.) I then ask how much more or less light is let in between each shutter speed on the dial. Except for one place where an insignificantly slight adjustment is made in the system, because the time increments between shutter speeds are doubled (or halved in the other direction), they immediately see that changing shutter speeds doubles or halves the amount of light let onto the film by a given source, all other things being equal. Letting light from a source go through an opening for one second, for example, allows twice as much light to come in as does allowing it to go through for half a second.

I then go through the same thing with regard to aperture setting. "Which lets in more light?" Obviously the larger opening will. Then I simply tell them that the aperture system is designed in such a way that the difference between adjacent apertures also lets in twice or half as much light, depending on the direction one is going. Then I ask them how much different the amount of light let in is if I were to increase the shutter speed by one click, and also increase the size of the aperture by one click. They see that is the same amount of light. I do it again, and again. All the changes make no change in the amount of light let in to the film. It does not take them long to ask their own question, which is some variation on "Then how do you choose which combination to use, if they all do the same thing?" I take that question to be one sign they are understanding what I have been teaching them so far. And that question allows me to go into the properties shutter speeds and apertures control in addition to **amount** of light.

Now one of the, I think unjustified, criticisms of the Socratic Method is that the questions are not so much logically leading in a way to give insight as they are psychologically merely prompting questions whose answers are obvious from verbal or psychological cues rather than from attention to the content. This may be what Rud means when he says that

Socrates begins by putting words in the mouth of the slave. I can discuss this criticism better with an example of my own from a criticism of my Socratic Method paper. There is a place in that paper where I report asking a group of third grade students how many **numerals** there are in our (Arabic, decimal) numbering system. At first the answer the class shouts out is "nine"; then someone says "ten" and there is some agreement with that. So I asked "Which is it, nine or ten," and they all yelled back "Ten!" I took that, whether mistakenly or not, as a sign that they were now including zero as a numeral that they had initially neglected to consider when they said "nine." My next question was "If we list the numerals in order, starting with zero, what will the list be?" I was taken to task by a critic of the method for essentially telling the students zero was a numeral. I don't think I did, because I was pretty certain their change from nine to ten implied they realized zero was a numeral to consider, but I can see how someone might think so. However, it is really an irrelevant criticism of the method, though it may be a justified criticism of this particular application of the method. Had I been more precise or more thorough, or had I thought it necessary, I would have asked something like "Why did you change from 'nine' to 'ten'?" or "What are the ten numerals?" If zero were still missing, I could have asked something that would have got them to realize it, by, say, holding up a closed fist and asking how many fingers I was 'holding up'. When they said "none," I could have asked how they might write "none" numerically. Or if I asked them what the lowest numeral was and they had said "One," I could have asked them whether there was not some number lower than one. I assume that somehow or other we could have easily got to zero, and to their recalling and recognizing it was the lowest numeral. At that point the class would have continued as it then did in the original.

As to the answers to the individual questions in a Socratic dialogue being prompted by cues other than logic, there are four responses I would make to dispute that. 1) You cannot get satisfactory answers to the later questions without first having gone through the earlier questions, so at least these later questions by themselves obviously do not contain the clues for answering them. 2) There are people who "get lost" and cannot answer the middle to later questions in a chain of questions that develops a line of reasoning. They cannot make connections as they go along. They report that they cannot "follow" the line of thought being developed. 3) When people do successfully follow a chain of questions, they usually report that suddenly they "see" how all this [whatever is being explained or taught] works. They display a "Eureka" experience that cannot be accounted for by their having answered easy individual questions whose answers were obvious by the way the questions were asked. And 4) one can often see students make insightful comments or ask penetrating questions that show they are starting to catch on to things greater or more encompassing than your individual questions have covered - comments and questions they could not make or ask at the beginning of the chain of questions.

Rick Garlikov

Notes

1. Rud at one point talks about Sophie Haroutunian-Gordon's operating from a "close" reading of the dialogues, and some of the comments he makes, especially about the geometric derivation Socrates goes through with the slave boy, indicate he and other critics may not be operating from that kind of reading. I would have thought any analysis would necessitate operating from a "close" reading, but perhaps that is neither required nor expected in "citation-based" or "reference-based" scholarship or analysis based on deconstruction. When Rud says, for example, "...Socrates makes Meno uncomfortable with his persistent questions" that strikes me as an odd description of the origin of discomfort of the participants in the dialogues. It is not that Socrates has persistent questions as a three-year-old might or as a heckler at a political rally might; and it is not that his questions are difficult or embarrassing. The crux of the discomfort in many of the dialogues is that Socrates' specific questions lead logically and, in a sense forcefully, to ideas the participants cannot, or do not want to have to, entertain or accept. And when Rud says things like "Socrates was devious and crafty" or that Socratically inspired teachers are good at playing some sort of "dialogue game," it seems to miss the point in the same way it would miss the point to analyze Einstein's work by saying only that Einstein was intelligent but unusual in his thinking, and that his theories are quite strange. Furthermore, since "devious" and "crafty" imply gamesmanship or trickery of some sort, their use paints a picture quite different from the picture I see of Socrates as I read the dialogues. I see the Socrates of the Euthyphro, Meno, *Republic*, Apology, and other dialogues I have read as insightful, brilliant, understanding, sincere, honest, and passionately opposed to loose and fallacious thinking. But I only came to see him that way after teaching modern versions of the same topics, and seeing that the views contemporary students and adults hold about those topics are not very different from the views portrayed by the Platonic participants, and that the same objections or arguments Socrates used, apply as well.

This article has been retrieved 1119 times since November 25, 1997.

Education Policy Analysis Archives

Volume 5 Number 21 November 25, 1997 ISSN 1068-2341

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass Glass@ASU.EDU.

College of Education

Arizona State University,Tempe AZ 85287-2411

Copyright 1997, the EDUCATION POLICY ANALYSIS ARCHIVES. Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

Academic Freedom, Promotion, Reappointment, Tenure And The Administrative Use of Student Evaluation of Faculty (SEF):

(Part IV)

Analysis and Implications of Views From the Court in Relation to Academic Freedom, Standards, and Quality Instruction

Robert E. Haskell [1]
University of New England

Abstract: In three previous papers, it was noted that while a controversial history of research on the reliability and validity of student evaluation of faculty (SEF) exists, it has not been typically viewed as an infringement on academic freedom. As a consequence, legal aspects of SEF are neither readily apparent, nor available. Moreover, SEF has not been generally seen as an infringement on, and detriment to, academic standards and quality instruction. The article is a review of SEF legal rulings analyzed in terms of their implications for academic freedom and quality of instruction in higher education.

Table of Contents

- **Overview of Academic Freedom**
    - Assumption # 3: Students as

- **Methods of Instruction, Grading, and Academic Freedom**
    - The Faculty Right to Select Teaching Methods
    - The Faculty Right to Assign Grades
- **The Courts, the University, the Faculty, and Setting Academic Standards**
    - Who Is the University That Sets Academic Standards?
    - SEF and Administrative Pressure to Maintain Enrollment Not Academic Standards
- **Academic Freedom and Assumptions Underlying Student Rights to SEF**
    - Assumption # 1: Student as Consumer and the University as Business
    - Assumption # 2: Higher Education as a Democracy
    - Qualified Evaluators
    - Assumption # 4: Student Learning as the Responsibility of the Faculty
- **SEF and Conflict of Interest in Relation to Student-Instructor Interface**
    - Curricula and Conflict of Interest
    - Economic Conflict of Interest
    - Release of SEF To Students and To the Public
- **SEF In a Larger Context**
    - The Global Exporting of SEF
    - Legal and Administrative Default-Enforcement of Faculty Allegiance
- **Conclusion**
- **References**

As indicated in previous papers (Haskell, 1997a, 1997b, 1997c), the history of legal rights demonstrates that issues not considered to have legal standing only come to have legal standing after a long process of advocacy, requiring the accumulation of data, coalescing judgements and arguments surrounding an issue. This series of papers on SEF is in the service of that process.

In the first article (Haskell, 1997a), I suggested that despite a history of conflicting research and views on the reliability and validity of student evaluation of faculty (SEF) used administratively, it has not been considered an infringement on academic freedom, and that to question the use of SEF is often seen as an attack on either student rights[2] or on evaluation of faculty performance in general.[3] Faculty and educational administrator views and surveys, along with other data, were reviewed as SEF is used in salary, promotion and tenure decisions. I proposed that the literature showed that SEF infringes on instructional responsibilities of faculty by providing a control mechanism over curriculum, course content, grading, standards, and teaching methodology.[4] I further proposed that SEF plays a significant role in current attacks on tenure, and that its role in a demographically diverse 21st century educational system has changed from its benign historical origins, concluding that contrary to current views SEF is a serious and virtually unrecognized infringement on academic freedom.

In a second article (Haskell, 1997b), I suggested that as a consequence of SEF not being viewed as infringing on academic freedom, its legal aspects have been neither readily apparent, nor available. Accordingly, as a legal category SEF has been virtually absent in compendia on higher education law. Legal rulings were abstracted and categorized from located SEF cases. In a third article (Haskell, 1977c), these legal rulings, their implications and assumptions in relation to their accuracy and psychometric validity, where SEF is integral to the denial of academic freedom, tenure, promotion, and reappointment, were reviewed along with the legal principles of Disparate Treatment, Disparate Impact, and the scientific Precautionary Principle in policy decision making.

This final paper will continue to examine legal rulings on SEF cases involving the denial of tenure, promotion, and reappointment decisions in relation to its implications and assumptions regarding academic freedom and quality of instruction. Finally, I would like to point out that the issues examined in this series of papers are not primarily concerned with individual faculty rights but with the implications of SEF when used for administrative purposes on academic freedom, educational quality, standards, and ultimately on the competence of graduates.[5]

Overview of Academic Freedom

Few higher educational issues are more important, controversial and ambiguous than the issue of academic freedom. Among many faculty and administrators, the concept of academic freedom, like the public's view of the First Amendment right to free speech, takes on a near *carte blanche* quality. It is therefore widely misunderstood.[6] Some restrictions on faculty speech in the classroom are, of course recognized. Just as it is generally understood that the public's right to free speech does not extend to the well-known limitation of loudly shouting "fire" in a crowded movie theater, so too faculty generally understand that academic freedom does not extend to classroom political commentary not directly related to the subject matter of their course. One related issue which apparently elicits more controversy than the concept of academic freedom itself is SEF. When I initially argued (Haskell, 1997a) that the administrative use of SEF is an abridgement of faculty's academic freedom, one critic (Theall, 1997) quickly objected, asserting that "Academic freedom has been defined in many ways, but never before in a way that suggests the construct (tradition? principle? tenet?) is vulnerable to the influence of student ratings." He was, of course, nearly correct (also see Haskell, 1977d).

In my initial paper, it was noted that SEF had neither legally, nor by published title, been identified as an infringement on academic freedom.[7] I was recently informed by a Canadian colleague, however, that there exists at least one early reference to SEF being an abridgement on academic freedom. This reference is a book chapter by Christopher K. Knapper (1977) entitled, "Teaching Evaluation and Academic Freedom." He opens his chapter, noting:

> A number of previous chapters have talked about faculty resistance to teaching evaluation, and have *hinted that the freedom of the individual professor to decide the content of his course and how it is to be taught may be infringed upon by the widespread use of formal evaluation procedures*. Some enthusiasts for student evaluation of instruction ignore this aspect of the question completely, others consider it in passing as a factor which must be taken into consideration in setting up evaluation programs. *Hardly any writers have tackled the issue head-on, to discuss the various ways in which academic freedom might be infringed* (p.198, italics added).

Academic freedom has many ambiguous facets; indeed, its core meaning may be dwarfed by its fuzzy outer periphery.

There is no shortage of past and current analyses of academic freedom (e.g., *Academe*, 1997; Dewey 1976 [1902]; Furedy, 1995; Lovejoy 1937; Menand 1993, 1996; Morrow and Sills 1968; Stichler, 1997). In a recent volume by Menand (1996), the closest and only allusion to SEF infringing on faculty behavior is found in Chapter Four. In discussing the formal and informal regulation of speech (i.e., certain viewpoints) Sunstein notes, "the evaluation of students and colleagues cannot occur without resort to [course] content, and it would be most surprising if viewpoint discrimination did not affect many evaluations" (p.106). Viewpoint discrimination will be examined below.

In 1973, the Commission on Academic Tenure in Higher Education (jointly sponsored by the AAUP and the AAC) made the following slightly more concrete recommendation relative to teaching incompetence: The commission believes that "adequate cause" in faculty dismissal proceedings should be restricted to (a) *demonstrated incompetence* and dishonesty in teaching and research (University of Michigan, 1994). [italics added]. The AAUP's Statement on Teaching Evaluation further suggests that "Casual procedures, a paucity of data, and unilateral judgments by department chairs and deans too often characterize the evaluation of teaching in American colleges and universities....A judicious evaluation of a college professor as teacher should include: (1) *an accurate factual description* of what an individual does as teacher" (AAUP Committee C , 1975). [italics added][8] The phrases, "*demonstrated incompetence*" by "*an accurate factual description* of what an individual does as teacher," have been central in the analysis of court rulings in this series of articles.

Further, most definitions, analyses and discussions of academic freedom are general and abstract. In addition to general statements about academic freedom including the freedom to teach, it is unclear what freedom to teach concretely means, with the exception that faculty are entitled to freedom of discussion and inquiry in their classrooms as long as they do not introduce controversial matter which has no relation to their subject. Even with this, what constitutes "controversial" and "no relation" is ambiguous in itself. This lack of concreteness in analyzing academic freedom is perhaps one reason why SEF has not been recognized as an abridgement of academic freedom. As this series of articles has demonstrated, SEF provides a concrete view of academic freedom from the teaching trenches, a view that reveals unrecognized infringements on academic freedom. A wider review of the concept of academic freedom is beyond the scope of this article.[9] Thus, except as it is concretely defined in this last of the series of articles on SEF, academic freedom must function as a "primitive" term.[10]

Methods of Instruction, Grading, and Academic Freedom

Most faculty (including myself until recently) seem to believe that academic freedom pertains not only to free speech in the classroom, but also to teaching methodology, grading, and assigning course work. Full-time teaching faculty traditionally have been and continue to be, by virtue of their disciplinary knowledge and daily classroom experience, the primary group that is fundamentally situated for defining standards in higher education. Indeed, it has reached near origin-myth proportions among faculty that the university is the faculty, and publically, administrators typically lip-sync this view. How have courts in relation to SEF cases in fact ruled on such thought-to-be time-honored prerogatives of the faculty---individually or collectively? And what are the implications for SEF, academic freedom, educational standards, and quality of instruction?

The Faculty Right to Select Teaching Methods[11]

According to the SEF cases analyzed in Haskell (1997b, 1997c), courts have ruled:

> Summary: Teaching method (72) is not a form of free speech, nor (73) covered under academic freedom, (74) except if noted in specific contractual faculty agreements. Numerous courts, (72) have separated faculty *speech* from *action* in the classroom, (2) have maintained that faculty can not disregard institutionally established curriculum content in the classroom; that the first amendment does not prevent a university from terminating an untenured faculty whose "pedagogical style and philosophy" does not conform to that of "the school's administration", (1) have further ruled that it is acceptable for untenured faculty to be terminated because of a refusal to lower their academic standards, (22) that a decision not to retain a non tenured instructor, even though based, in part, upon SEF that express disapproval of the faculty teaching methods does not violate a faculty's First Amendment right to academic freedom.[12]

Thus, perhaps one of the most widely misunderstood aspects of academic freedom is the faculty right to decide their teaching methods, i.e., what faculty *do* in the classroom not what faculty *say*. The Supreme Court has stated: "Any inhibition of freedom of thought, and of *action* upon thought in the case of teachers brings the safeguards of those amendments [First and Fourteenth] vividly into operation" (Shelton v. Tucker, 1960, italics added). The term "action" could be construed as relating to teaching method.

In *Carley v. Arizona Board of Regents* (1987), the faculty member contended that teaching methodology was part of his academic freedom right (see endnote # 9). The court clearly said it was not, citing numerous other cases and legal principles to support their denial. Carley cited several rulings in support of his position, but the court disagreed with them, saying that the cases he cited involved conduct more closely resembling speech than teaching method. For example, in *State Board for Community Colleges and Occupational Education v. Olson*, (1984), the court stated that the "principle [of academic freedom] finds its source in the belief that teachers should be free to engage in the exchange of diverse *ideas* on controversial topics..." (p. 437, italics added). This case involved the cancellation of a student newspaper that was part of a class. The court found that canceling the student newspaper did not "abridge the constitutionally protected aspect of [her] teaching function," as Olson was still free to utilize other instructional methods for "presentation of the idea-content of her journalism courses..." (p.1101). In this context, teaching method was not protected as is speech.[13]

The court also cited *Clark v. Holmes* 9474 F.2d at p. 931(1972), where the court upheld the nonrenewal of a non tenured instructor for reasons related to the structure of his course content. The court stated "We do not conceive academic freedom to be a license for uncontrolled expression at variance with established *curricular contents* and internally destructive of the proper functioning of the institution" italics added). The court rejected the claim, based on the same distinction articulated in Lovelace involving homework assignments, course standards, and the distinction between speech on the one hand and teaching methods on the other. In *Lovelace v. Southeastern Massachusetts University* (see #9), the court ruled that Lovelace's nonrenewal was because of what he *did*, not what he *said*. Specifically, according to the court, student complaints about his grading policy had nothing to do with his speech. As the court noted, "*Matters related to grading policies, course content and homework load are policy matters for the university*" (italics added). Exactly who constitutes the university will be addressed below. Similarly, the court ruled that Carley's complaint did not involve speech.

In the Carley case, the court also cited *Hetrick v. Martin* (1973), in which a state university declined to renew the appointment of a non tenured faculty member due to disapproval of her "*pedagogical attitude*" (italics added), as evidenced by teaching styles and techniques. The court expressly refused to recognize teaching methods as protected speech, holding: Whatever may be the ultimate scope of the amorphous 'academic freedom' guarantee to our nation's teachers and students...it does not encompass the right of a non tenured teacher to have her teaching style insulated from review." The court finally cited other cases which ruled that teaching methods do not generally fall under the rubric of academic freedom.[14] Indeed faculty challenges to institutional denial of tenure decisions for reasons relating to instructional methods, course content, and grading policies have generally been unsuccessful.

According to Copeland and Murry (1996), the courts have repeatedly ruled that colleges and universities have broad control over course content, homework, and grading policies, and over pedagogical methods. Another legal writer concludes (Weeks, 1988), "that challenges to the use of student evaluations based on a claim to academic freedom will not be sustained as long as those evaluations focus on teaching method, classroom presentation, and general teaching skills" (p. 6). When on the basis of student complaints about (reasonable) teaching methods and institutions' denial of tenure on the basis of those student complaints, and when the courts deny that teaching methods are protected by academic freedom, there exist a number of serious problems, educational problems that have not been adequately addressed.

The historical distinction in educational cases between *speech* on the one hand and *action* or *methods* on the other hand appears to engender at least one basic false assumption, and a host of pedagogical implications, including implications for academic freedom. One false assumption is this: That course content and teaching method are often two inherently distinct areas. The distinction between speech and action or content, however, is analogous to that of *content* versus *form* discussion in the humanities and elsewhere (e.g., in art, literary criticism, philosophy). It is generally accepted that in many situations that content and form can not be clearly separated.

Occasionally, the courts appear to recognize this false dichotomy. For example, in the Carley case, the court cited *Kingsville Independent School District v. Cooper*, (1980), wherein the court reviewed a history teacher's presentation of post-Civil War Reconstruction using a role-playing technique which evoked strong student feelings on racial issues. The school board declined to renew her teaching contract because there had been complaints by parents about her instructional method. The court noted that like Olson, the case involved the discussion of controversial topics and the presentation of controversial course materials. Unlike in Olsen, however, the court defined teaching methods as speech and found that the speech *was protected* and could not be used as a basis for non renewal of the teacher's contract. However, in the Carley case, the court found that in his commercial art course the requiring of business values, e.g., of being prompt, self-reliant along with his method of instruction being demanding was not academically protected. Other than the fact that these rulings came from two different courts, the distinction between these two cases appears to be one without an *instructionally* pragmatic difference. In the Kingsville case, for example, the school prohibiting role-playing as a teaching method was found to be a violation of speech. In point of fact, then, the court *in effect* did rule in favor of teaching method. It is difficult to understand why this method was protected under free speech since the same historical content could be taught without using role-playing as a teaching method.

It is unclear what the difference is between the Kingsville case involving role playing, and the Olson case, in which the court found that canceling the student newspaper did not abridge the constitutionally protected aspect of [her] teaching function because Olson was still free to utilize other instructional means for "presentation of the idea-content of her journalism courses..." (p.1101)? So, too, could the teacher in the Kingsville case utilize other instructional means for presentation of the idea-content of her history course. Given the goals of instruction, it is not easy to see how Carley could have used other instructional means to deliver what he saw as the appropriate (business) content of his course. As I noted in Haskell (1997c), clearly the courts' logic in such cases becomes

> unwieldy, not just to the nonlegal scholar, but apparently to the Courts as well....To the layman, legal rulings regarding SEF are a veritable thicket, often seeming that the use of context to differentiate one apparently similar case from another functions as a kind of ad hoc carte blanche to justify preconceptions and positions.

Finally, at least in certain realms adjudged by the U.S. Supreme Court, the distinction between speech and action is often considered a distinction without a difference. For example, in 1989 the court ruled that flag desecration (flag burning) could be a form of political expression and, as such, would be protected under the First Amendment, which specifically prohibits Congress and the states from making any laws to abridge such freedom of speech/action. Granted, flag burning is judged under a more stringent legal interpretation of free speech because it is considered to be in the *political* arena, and therefore more important than other situations, hence the denial of the distinction between speech and action. But what is more important than the means (action) by which we educate student minds?

Three pedagogical examples may help clarify the fusion of course *content* with instructional *methods*: It is generally accepted in the small group dynamics literature that if the goal is teaching students to function in small group conditions (as opposed to simply learning *about* small groups), that one of the most effective *methods* (if not the most effective method) is having students actually function as a group (often known as an experiential method, where students learn by experiencing the group processes). What if students complained about this method? Could---and should---administration, fearing the loss of tuition dollars, legally prohibit this teaching method? Would the courts rule on a suit as they did in the Carley and Lovelace cases? Or would they rule as they did in Kingsville? There seems to be no principled way of knowing. And what if administration decided that the experiential method was inappropriate because the course limits enrollment to only 15 students, or because the experiential method was otherwise pedagogically inappropriate?[15]

The second example involves courses, where students are required to spend time in actual work-type situations like internships and practica relating to their major. Certainly internships fuse content with method. If students complained, could such methods be ruled inappropriate? In these instructional situations, students seldom complain, however, because like the small group experience and unlike a research-based course, many like such concrete (and so-called "real world")---yet unrepresentative experiences. The third example involves the question: is teaching course content, based on findings from the research literature in a content area, a teaching method? In this situation, course content fuses with instructional method. In teaching psychology, to non majors at least, many students do not like a research-based approach to the subject. For example, many students come into class with a pop psychology belief system which typically means a clinical orientation and they want interesting anecdotal illustrations. In principle, then, it is faculty who would seem to be the best judge of what teaching methods are appropriate for their areas of expertise.[16]

The Faculty Right to Assign Grades

Just as most faculty seem to believe that academic freedom pertains not only to free speech in the classroom, but also to teaching methodology and required course work, it is also generally believed that faculty are the final authority for assigning grades to student work. Even this traditionally viewed sacrosanct prerogative, however, is dependent on a number of contextual conditions.

In the SEF case, *Lovelace v. Southeastern Massachusetts University* (see endnote #9), the court ruled that the non renewal of Lovelace's contract was *because of what he did, not because of what he said.* Specifically, the court ruled that student complaints about his grading policy had nothing to do with his speech, noting, "*Matters related to grading policies...are policy matters for the university*" (p.424, italics added). It is common practice for institutions who have accepted AAUP guidelines to develop grade-change policies and a set of explicit procedures, stating that state prior to any change of grade assigned by a faculty that the institution and/or a faculty committee shall notify the faculty member of any such change and the reason for the change. The crucial issue of who constitutes the university will be addressed in detail below.

Although many cases support institutional authority over faculty instructional activities, including grading, faculty academic freedom in matters of grading can prevail over institutional authority. In a case that is often cited, *Parate v. Isibor*, 868 F.2d 821 (6th Cir. 1989), basing their decision on the First Amendment, the court limited the deference traditionally accorded administrative decisions about grading of students (Kaplin and Lee, 1995, see section 3.7.2. Academic Freedom in the Classroom 311). The dean of the school in which the faculty was a non tenured professor ordered the faculty to change a final grade of one of his students. The faculty member argued that his dismissal was in retaliation for his lack of cooperation regarding the grade change and therefore violated his First Amendment academic freedom. Relying on the Free Speech Clause, a court agreed saying that

> [B]ecause the assignment of a letter grade is symbolic communication intended to send a specific message to the student, the individual professor's communicative act is entitled to some measure of First Amendment protection.

The court further reasoned:

> [T]he professor's evaluation of her students and assignment of their grades is central to the professor's teaching method.... Although the individual professor does not escape the reasonable review of university officials in the assignment of grades, she should remain free to decide, according to her own professional

> Judgment, what grades to assign and what grades not to assign.... Thus, the individual professor may not be compelled, by university officials, *to* change a grade that the professor previously assigned to her student. Because the individual professor's assignment of a letter *grade is protected speech*, the university officials' action to compel the professor to alter that grade would severely burden a protected activity [868 F.2d at 828].

Thus, the Dean's act of ordering the faculty to change the grade, contrary to the faculty's professional judgment, violated his First Amendment right.[17]

A further significant aspect of the court's ruling, however, is this: had university administrators changed the student's grade themselves, the Dean's action would not have violated the faculty's First Amendment rights.[18] As Kaplin and Lee (1995) point out, "The protection that Parate accords to faculty grading and teaching methods is therefore quite narrow—more symbolic than real, perhaps, but nonetheless an important step away from the deference normally paid institutions in these matters."[19]

The Courts, The University, The Faculty, and Setting Academic Standards[20]

From the legal cases involving SEF reviewed in this series of papers, the courts have consistently ruled as follows:

> Summary: From the cases analyzed, the courts have clearly said (68) universities must be allowed to set standards, including (69) course content, (70) homework load, and (71) grading policy

Indeed, with few exceptions, the courts have traditionally taken a hands-off approach to academic matters. As noted in the often cited *Regents of the University of California v. Bakke* (1978) case, which is generally seen in the literature as paradigmatic, the court ruled "the four essential freedoms of a university are "to determine for itself" on academic grounds who may teach, what may be taught, how it shall be taught, and who may be admitted to study." So the courts have most always maintained that it is not in their purview to set academic standards. The pertinent question here that is seldom addressed explicitly and categorically is: who exactly constitutes "the university?" Is it faculty? Is it administration? Or is it a cooperative/compromise faculty/administrative set of policies? The question of who constitutes the university, both legally and normatively, is central to understanding court rulings.

Who is the University That Sets Academic Standards?

Traditionally---at least in time-honored principle---the answer has been that it is the purview of faculty to set academic standards, including curricula, course content, and expected level of student performance. Legally, however, it has always been clear that the university boards of trustee and Regents are the final legal arbiters of academic standards. In fact, the latter *are* the university. But as one classic statement on academic freedom clearly points out (Morrow, 1968),

> This latter usage is clearly distinct and derivative; for such corporate autonomy derives its justification ultimately from the services performed by the scholars whose activity it exists to foster and protect, while, on the other hand, the freedom of the individual scholar often requires protection from the pressures of his own institution, as well as from outside forces (p.4).

Nevertheless, legally, faculty set standards only in so far as they are accepted by specific contractual agreements with these bodies through administrative procedures. As Kaplin notes, "The four essential freedoms concept would apply to both policies developed by administrators and policies developed by faculty if the policies are adopted by the university as university policy. Technically, the four freedoms (who may teach, what may be taught, how it shall be taught, and who may be admitted to study) are attached *to institutional policy not to individual administrators* (Personal communication, April 1997, italics added).[21]

A cardinal problem, however---at least in relation to SEF rulings---is that the courts often accept certain decisions by individual administrators *as if* they were the formally delegated authority of the boards of trustees and Regents---even in the formal absence of a particular regulation or policy regarding a decision. In short, the courts tend to accept an administrator's decision *as if* it were based on formal institutional policy. This is important to recognize. For example, as will be noted below, few universities have formal policies about teaching methods or grading. Courts, then, often use the term "university" (1) *as if* it means any pronouncement by a university regardless of the existence of a formal policy, (2) or any pronouncement by a single administrator, e.g., a Dean---perhaps assuming the pronouncement to be a formal policy. For example, In *Lovelace v. Southeastern Massachusetts University* (1986), the court ruled, "The first amendment does not prevent a university from terminating an untenured faculty whose pedagogical style and philosophy does not conform to those of the school's administration." Pedagogical style and philosophy more often than not includes grading.[22] A collective faculty agreement (e.g., a union contract, agreements in a faculty handbook, etc.) notwithstanding then, courts tend to uphold individual administrator's decisions on course content and teaching methods even when course content and teaching methods are not explicitly a part of written policy.[23]

Thus a major variable in any legal ruling is how an issue is addressed in faculty union contract agreements and faculty handbooks. While in the Northern Arizona University case the court ruled that teaching style is "not a form of speech protected under the First Amendment" (Heller, 1986), the arbitration board ruling on the use of SEF at the University of Guam was based on a violation of the union contract (Blum, 1990). One wonders what the Arizona ruling would have been if teaching methodology was explicitly stated in a faculty contract. Other union contracts similarly prohibit the use of SEF for administrative evaluation purposes.[24] While policy on course content and teaching method may be included in some faculty agreements, I am aware of only one agreement that includes statements on pedagogical style and philosophy.[25] Article V of the Vermont colleges statement on academic freedom, for example, states,

> It is the Policy of the Vermont State Colleges to maintain and encourage full freedom of inquiry, teaching and research. Academic Freedom implies not only the unconditional freedom of discussion in the classroom, *but also the absence of unreasonable restrictions upon the classroom instructor's methods* (italics added).[26]

This, perhaps singular---at least rare--policy, is crucial for faculty if academic standards are to remain their prerogative.

It is clear that if faculty want to maintain this prerogative, they must explicitly have in their agreements and handbooks polices on teaching methodology, grading, and SEF.[27] At least to the extent that standards are not collectively set by faculty, educational policies such as classes being composed of students with widely---and inappropriately---varying ability, class size, and grading standards are outside faculty purview, yet all of them have impact on SEF. Even so, institutions do not legally have to agree to faculty-developed standards. As Stone (1995)observes:

> At one regional institution, faculty surveys repeatedly reported widespread concern about institutional reliance on student ratings of instruction as a basis for faculty evaluation. The faculty senate of the institution tried repeatedly to address the problem, but the solutions desired by the faculty were administratively rejected or shunted aside, in part, because of concerns about student satisfaction. Thus, the administrative view prevailed.

Such overruling of a faculty body is not infrequent.[28]

Ideally, for the courts to take a hands-off approach, refusing to second guess academic matters is the prudent course. But the world is not ideal, as demonstrated by the courts more hands-on approach in discrimination, e.g., Disparate Treatment and Disparate Impact cases (see Haskell, 1997c). In addition, the world is not ideal in other respects as well, especially pertaining to SEF. By assuming validity of SEF data and "good faith" on the part of university administrators, the courts, by default or in a *de facto* manner, are setting academic standards. The question now is, is the "good faith" attributed to university boards of trustees and administrators justified?

SEF and Administrative Pressure to Maintain Enrollment Not Academic Standards[29]

Clearly, administrators are tending to assign increasing importance to SEF for tenure and promotion decisions (see Haskell, 1977a). The question is: why? The obvious answer is that they are concerned with the quality of education. This assumes, however, that administrators are more concerned about quality education and are acting more in the best interest of students than are faculty.[30] Understandably, the courts are apparently not aware of numerous changes and pressures within higher education. For example, with increased competition for students, institutions have become increasingly concerned with maintaining student enrollment and tuition monies as well as to reduce the number of tenured faculty. As noted below, consumerism is *in*, academic quality is *out*. This view is not a politically radical one; nor is it a well-kept secret.

It seems reasonable to conclude that courts have in effect not taken a hands off approach, but have been setting academic standards (albeit perhaps unintentionally)---by apparently assuming that institutional administrators' primary goal is quality education. As indicated above, courts have tended to accept administrative (subjective) judgements if they appear "sincere," grounded on some evidentiary basis, made on the "vigor and variety of student criticisms," and "not arbitrary or capricious and were exercised honestly upon due consideration." These rulings make a number of naive assumptions, assumptions that at one time may have been relatively valid. As I previously observed (Haskell, 1997c), "The courts continue to assume a kind of pre 1960s academic Camelot. If such a round table of academic knights ever did historically exist or was merely mythical, it certainly now exists only in myth." As Copeland and Murry (1996) have noted, "the judiciary has tended to act as if colleges and universities could be trusted to act in good faith" (p.246). Courts should no longer assume altruistic institutional motives that were perhaps true prior to the 1960's.

In *Robert Kramer v. The President of the University of British Columbia* (1992), for example, it was noted by one student "that some of the unhappiness came from the fact that the levels of Japanese language ability were badly divided; some found it easy, others very hard" (p.12). It was further noted that the department head viewed Kramer's course evaluations "with some alarm" and that a number of students had stated that Dr. Kramer's teaching would cause them to stay away from the Asian Studies department. Since the course was the general introduction to the subject, such negative comments were of great concern to the department head. In another report of a faculty dismissal, involving low student evaluation of a faculty member because of his class standards and requirements, administration was quoted as saying, "We're an open-admission university. A large fraction of the class was completely unable to compete" (Magner, 1995).

In another case (*William Sypher v. Vermont State Colleges Faculty Federation*, 1982), the court seemed to accept that whatever level of student is enrolled in a class, the instructor is obligated to teach even though standards may be lowered. Responding to being denied reappointment, and in defense of his student rating level, Sypher wrote a letter to the Dean in which he said, "It is certainly distressing when very good is not good enough, especially at a college with a modestly-talented student body that often discourages efforts at subtlety, wit and deeper penetration of subjects." The Board responded to his letter saying, "*other actions and statements* by Grievant constituted legitimate reasons for not retaining him. In a May, 1980 letter to [the Dean], Grievant *expressed his contempt* for Castleton students" (p.135), concluding, "*Accordingly*, we find credible the College's contention that Grievant was not reappointed because of his teaching effectiveness" (p.135, italics added).[31] SEF, then, plays an important role in maintaining student enrollment.[32]

SEF also plays a role in the continuing attempts to reduce the number of tenured faculty on a campus. As I explained in a previous paper (Haskell, 1997a)

> What is not widely understood is that SEF is often a kind of Trojan Horse in the battle against tenure and academic freedom. It often becomes a stealth mechanism by which to covertly abrogate both tenure and academic freedom.

In a rare published recognition of the stealth role of SEF in abrogating tenure, Knapper (1977) suggested

> That complaints about teaching are commonly cited as a cause for dismissal is, in a rather perverse way, encouraging, if it reflects an increasingly important status assigned to this academic function. On the other hand, it may be that student complaints about teaching are the symptom, not the cause, of the trouble, or that teaching difficulties are cited as the excuse, but do not really constitute the main reason, for dismissal (p.199, italics added).[33]

As already noted, courts do not trust that universities will act in good faith with regard to discrimination cases. They do, however, tend to assume good faith with regard to pedagogical methods, grading, and other standards.[34]

Considering the literature on the lowering of standards in higher education, then, it would seem that the courts should no more automatically assume good faith in educational matters than they do in matters of discrimination. In assuming good faith the courts are setting academic standards by default. But this is not likely the end of it: Pressures to grade easily, to lower standards in terms of content and requirements, and to pass students who have not earned passing grades takes on a different meaning in light of recent court cases by students charging the institution with not teaching them to an adequate level of competency (American Psychological Association Monitor, 1994; Chronicle of Higher Education, 1996).

Given the cases presented here demonstrating the SEF-driven administrative pressure on faculty to lower their classroom standards, the consequences of such practices already have reached the courts. Increasingly, students are suing universities, claiming that the education they received was poor or failed to live up to promises made in course catalogues. As higher education focuses increasingly on vocational "education," such cases may also increase. Institutions can not have it both ways: to pressure faculty to grade more easily, on the one hand, and ensure student competency, on the other. They do not go together.

Academic Freedom and Assumptions Underlying Student Rights to SEF

There are a number of assumptions which undergird the new right of students to evaluate faculty and its use in administrative decisions for reappointment, promotion and tenure, assumptions which have been both explicit and implicit in the court rulings on SEF. Four of the more significant assumptions are: (1): Student as Consumer, (2), Higher Education as a Democracy, (3) Students as Qualified Evaluators, and (4) Student Learning as the Responsibility of the Faculty.

Assumption # 1: Student as Consumer and the University as Business

SEF has come to be seen as a right. There is a pervasive business metaphor of consumerism in higher education (an assumption that I briefly address in Haskell, 1997a). The university considered as a business carries with it the attendant and associated ideas of students as *consumers* in an *educational marketplace*. As the well-known scholar David Reisman (1981) noted years ago in this regard, "This shift from academic merit to student consumerism is one of the two greatest reversals of direction in all the history of American higher education; the other being the replacement of the classical college by the modern university a century ago" (p.xi). From this model, it follows that students are consumers of instruction and therefore have a right to evaluate and influence instruction. To question the-student-as-consumer right is difficult. As I noted in a previous paper (Haskell, 1997a)

> While it is difficult enough to deal with political, ideological and economic pressures, dealing with consumer pressures has become nearly impossible. If denying fiscal efficiency is viewed as unreasonable, irresponsible, and even irrational, to deny "consumer's" their demands is viewed as undemocratic and downright mean spirited.

Some state-supported schools advertise themselves as a business.[35] The fact is, a university is not---or should not be---like a business. The business metaphor is an inappropriate one. And at least one court has recognized the inappropriateness of this metaphor.[36]

As McMurtry (1991) Damron (1995) and others note, the metaphor of consumerism itself is based on a number of incorrect and opposing assumptions.[37] Stone (1995) suggests that

> Higher education makes a very great mistake if it permits its primary mission to become one of serving student "customers." Treating students as customers means shaping services to their taste. It also implies that students are entitled to use or waste the services as they see fit. Thus judging by enrollment patterns, students find trivial courses of study, inflated grades, and mediocre standards quite acceptable. If this were not the case, surely there would have long ago been a tidal wave of student protest. Of course, reality is that student protest about such matters is utterly unknown. Tomorrow, when they are alumni and taxpayers, today's students will be vitally interested in academic standards and efficient use of educational opportunities. Today, however, the top priority of most students is to get through college with the highest grades and least amount of time, effort, and inconvenience.

Stone recognizes that "student ratings of instruction can serve as valuable feedback to an instructor about student preferences, but there is good reason to suspect that using them as a basis for administrative decisions on promotion, tenure, and merit pay has been a major contributor to the academic decline and devaluation of the past twenty-five or so years." There are further assumptions underlying the consumer metaphor. As Damron correctly observes,

> The concept of "student as consumer" begs a number of substantive issues, not the least of which is who pays for the services rendered to students and who benefits from them. In most universities and colleges, the lion's share (well over 80%) of the cost of post-secondary education is paid by tax payers. As McCabe (1981) has noted, virtually all college students are on sizable public scholarships. The remainder is defrayed by private stipends and scholarships, low interest government loans, or students and their parents. Thus, in the vast majority of cases, tax payers, governments, parents, and contributors to scholarship funds have an important stake in the services provided to students and are rightly conceived of as consumers of college teaching and its products. So too are the employers who demand and support professional programmes and routinely hire their graduates. Similarly, universities that accept college transferees are also consumers of college teaching insofar as they accept the curricula of colleges as reasonably equivalent to their own and grant transfer credits to those entering their degree programmes. If consumer satisfaction is to serve as the criterion of effective teaching, the satisfaction of all of the above parties must be assessed. Given the concerns about the quality of secondary and post-secondary education recently expressed by these groups, it is unlikely that advocates of the "student as consumer" will find this palatable.

Clearly the simple student-as-consumer metaphor is inappropriate.[38] Thus the "consumer" of higher education is in fact a wide constituency of groups distributed in both space and time. The metaphor of student as consumer is more appropriately replaced by the metaphor of student as worker or as apprentice.

The consumer metaphor and its implementation via SEF used for administrative purposes constitutes the newest threat to academic freedom and instructional quality.[39] McMurtry (1991) has noted that, education has always been subject to external pressures whose purpose is to subordinate it to vested interests of various kinds, whether it is slave-holding oligarchies, theocratic states, political parties or the prevailing dogmas of collective beliefs. I suggest that the difference today is that threats to academic freedom come from within: the consumer-student.

Assumption # 2: Higher Education as a Democracy

Closely related to the business metaphor of consumer is the tandem political concept of democracy, and like the former, the latter undergirds the administrative use of SEF. In a democracy it is always tempting to transfer this political principle to just about every realm of life, including higher education. So pervasive are the metaphors of student-as-consumer and higher-education-as-a-political-democracy, that to question one is to almost automatically question the other. And just as to question SEF is seen as undemocratic, to question the democracy metaphor is seen as totalitarian. Nevertheless, both metaphors are inappropriate when applied to higher education, culminating in counterproductive outcomes.

As the cum liberal, cum conservative historian (depending on who is doing the evaluating), Christopher Lasch (1979) has observed,[40]

> The democratization of education has accomplished little…It has neither improved popular understanding of modern society, raised the quality of popular culture, nor reduced the gap between wealth and poverty…On the other hand, it has contributed to the decline of critical thought and the erosion of intellectual standards, forcing us to consider the possibility that mass education, as conservatives have argued all along is intrinsically incompatible with the maintenance of educational quality (p.222).

As I suggested in a previous paper (Haskell, 1997a), SEF is a major factor in grade inflation and the erosion of academic standards.[41] On some campuses, the situation has reached the notice of accrediting agencies with the grade inflation noted in their reports.[42]

Most students do understand the ensuing consequences of the consumer and democratic metaphors underlying SEF. A glance at articles from online student newspapers reveals strong sentiments against what some students consider the erosion of standards created by SEF.[43] Some students are thus quite aware of the effects of SEF on their education. SEF has become such a matter of amusement---when they are not detrimental to ones career---that a recent article in a prestigious psychological journal that is not given to publishing such articles has published an apparently serious piece on "How to improve your teaching evaluations without improving your teaching" (Neath, 1996). It is time to get beyond the ideologies of student consumerism and democracy and begin dealing with the educational consequences.

Assumption # 3: Students as Qualified Evaluators

Validity of assessing teaching effectiveness assumes qualified assessors. In both the consumer and democratic metaphors, it is assumed that students are qualified to judge and assess. The fact is that even under ideal conditions most students are not qualified to judge faculty teaching effectiveness.[44] Ideally they may be able to judge certain aspects of teaching, e.g., clarity of presentation, instructor being organized, interest level, etc. In addition, the assumption of student as qualified evaluator of teaching effectiveness in turn subsumes numerous other assumptions. These assumptions include appropriate (a) maturity level (b) ability level, and (c) good faith motivation. A most cursory of glances at professional articles, reports, periodical media, books, and educational world-wide-web Internet sites yields an abundance of documentation on grade inflation, low SAT scores, lowered academic preparation of students, and lowered admissions requirements, with a consequent increase of remedial college courses (Adelman, 1996;[45] American Federation of Teachers, 1996; Bauer, 1996; Blum, 1992; Brimelow, 1996; *Chronicle of Higher Education*, 1997; *Chronicle of Higher Education*, 1991; Fighting grade inflation, 1994; Gordon, Hartigan and

Muttalib, 1996; Goldman, 1993; Gose, 1997; Guernsey, 1996; Hertling 1996; Kolevzon 1981; Lambert, 1993; Lasch, 1979; Leo, 1996; Sacks, 1996; Simon, 1996; Stedman, 1996; Stone, 1995; Summerville, et al, 1990; Walker, 1992). Thus, acceptance of the above assumptions seems highly questionable. Indicating and documenting just how widely acknowledged the recognition of the inappropriate acceptance of these assumptions is and their effects on higher education, it has reached comic strip proportions as the frequent subject of the popular Doonsbury comic strip.[46]

Student level of (1) preparedness, (2) ability level as measured by most any national test, (3) expectations about learning, (4) motivation level, and (5) hours spent studying have all been in decline for years, yet at the same time a sense of entitlement, motivation, and the average grade has risen from a C perhaps C+ to a B, perhaps a B+.[47] SEF contributes more than its share to this state of affairs (see Haskell, 1997a). The Higher Education Research Institute recently released the National Norms for the freshman class of 1995. The survey was completed by 323,791 freshmen entering 641 two-and four-year colleges and universities. Among other characteristics, the survey showed that students tend to be increasingly disengaged academically. Students are spending less time studying and doing homework, with the percent reporting six hours or more per week dropping from 43.7 percent in 1987 to 35.0 percent in 1995; spend less time talking with teachers outside of class (47.0 percent reporting one or more hours per week, compared with 62.0 percent in 1989), and the highest percentage of students ever (33.9 percent) reporting being frequently bored in their classes.[48]

As demonstrated in the legal opinions on SEF, however, these realities seem not to have reached the court. Moreover, as Abrami (1989) observes, there is little to no rigorous research demonstrating the complex network of relationships existing between student impressions about the processes of instruction and the impact of those processes on student cognition and their affective responses to it. Despite this data, many researchers still maintain that SEF data is valid and appropriate for use in tenure, promotion, and reappointment decisions (given that certain adjustments are made in the SEF form and analysis of the data).[49]

Assumption # 4: Student Learning as the Responsibility of the Faculty

A further assumption underlying SEF that seems to be upheld mostly by default in court rulings, and by explicit educational philosophy in much of higher education, is that responsibility for student learning lies with faculty. SEF indeed holds faculties largely responsible for most of student learning. Granted, to some undetermined degree this pedagogical value can be justified. After all, the primary (or at least public) purpose of SEF is to attempt to establish "teaching effectiveness" ---which in fact means being responsible for student learning. Now, given (a) the lowering of admission standards, which has lead to (b) the admission of students that would historically not have been admitted, (c) the above suggested inappropriate maturity level, (c) inappropriate expectations, (d) inappropriate study time, and (f) lack of good-faith motivation by large numbers of students, how reasonable is it to hold faculty responsible for student learning?[50]

Amongst many faculty and administrators, the assumption seems to be that large numbers of students are unable to learn appropriately and require "inordinate" assistance from faculty. It is said that they are young and immature and thus require considerable nurturing. This view is perhaps appropriate for high school students, but not for college level students, at least to the degree it is considered to be required. Indeed, some courts have indicated that there is a maturity line between secondary and post secondary expectations and regulation of students. For example, in *Lansdale v. Tyler Junior College* (1972), considering the applicability to post secondary education of a prior precedent permitting high schools to regulate the length of students' hair, the court refused to extend the precedent. As one scholar explained (Kaplin and Lee, 1995):

> The college campus marks the appropriate boundary where the public institution can no longer assert that the

> regulation of . . . [hair length] is reasonably related to the fostering or encouraging of education. There are a number of factors which support the proposition that the point between high school and college is the place where the line should be drawn.... That place is the point in the student's process of maturity where he usually comes within the ambit of the Twenty-Sixth Amendment and the Selective Service Act, where he often leaves home for dormitory life, and where the educational institution ceases to deal with him through parents and guardians.[51]

In a more recent case involving community college students (*DiBona v. Matthews*, 269 Cal. Rptr. 882 (Cal. Ct. App. 1990), a California Court of Appeal ruled that administrators violated a teacher's free speech rights by canceling a controversial play production from a drama class. Distinguishing the case before from those involving minors in elementary and secondary schools, the court ruled that the college could not cancel the drama class solely because of the vulgar language in the play (Kaplin and Lee, 1995).

Each year state and federal courts render decisions in numerous cases involving both elementary, secondary, and post secondary education and have recognized that these precedents cannot be uncritically applied to each other. As Kaplin and Lee (1995) observe

> The majority holds today that as a matter of law the college campus is the line of demarcation where the weight of the student's maturity, as compared with the institution's modified role in his education, tips the scales in favor of the individual and marks the boundary of the area within which a student's hirsute adornment becomes constitutionally irrelevant to the pursuit of educational activities (p.13).

Unlike the courts, many faculty and administrators have been increasingly blurring this educational line---by design and, in effect---transforming college level men and women into high school students. I suggest this has been accomplished in part by lowering admission requirements, academic standards, and expectations. In addition, I suggest that many faculty, administrators, and educational theorists have been uncritically transferring research findings and teaching methods originally designed for elementary and secondary levels to higher education.

What most teaching methods and educational philosophy that are transferred from research on elementary and high school learning environments do---in addition to lowering the content level of the subject---is simply to use the college classroom for work that students should be doing outside of class: practice, memorizing, reviewing and other work. This is an additional way that higher education standards are lowered (See Chatterley and Peck's [1995] "We're crippling our kids with kindness!!" in the *Journal of Mathematical Behavior*, as an example of this on the elementary level.)[52]

SEF and Conflict of Interest in Relation to Student-Instructor Interface

SEF used administratively sets up inherent conflicts of interest between student and instructor. Perhaps first and foremost, to have students evaluate faculty for administrative purposes place faculty in an educational, a "political," as well as a potential economically vested interest relationship to students. These are not appropriate roles for educators to be forced into. When political and economic pressures impact instruction, clearly, the education of students is in danger of being degraded.

Curricula and Conflict of Interest

SEF not only can affect individual faculty, it can affect curricula as well. A faculty member at Wichita State University (Goldman, 1993) notes that in a thirty-faculty education department, responsible for certification of teachers, six faculty have been hired in the past 25 years as assistant professors to teach Foundations of Education. All faculty were apparently well qualified, receiving their doctorates from excellent universities. Only one of these faculty has been awarded tenure; none was promoted. According to Goldman, the reason for this was student evaluations. In general, as the data show, required courses hold less interest and receive lower evaluations than elective courses. Moreover, students seem to especially dislike the course in educational foundations. In addition, students who are drawn to become teachers are concrete-sequential, and are less interested in the abstract and theoretical content of the foundations of education course.[53] This leads the faculty who teach the foundational course to receive lower student evaluations than other education faculty. Because student evaluations are often---at least in effect---the primary, if not the only, gauge of teaching quality, and since teaching evaluation usually out-ranks research and scholarly productivity on most campuses, when tenure, promotion, and salary increases are awarded, these rewards will not be evenly distributed to faculty who teach the foundations and educational psychology courses. Rewards will accrue to the concrete-oriented methods faculty whose courses will further intensify the concrete orientation of teacher preparation. This in turn can then lead to a downward spiral in teacher preparation.

Economic Conflict of Interest

Indeed, the pressure for some faculty to conform to SEF is great. In one case (*King's College v. Anne S. De Fabry*, 1983) a faculty member requested of some students that they write a letter of recommendation for her application for promotion to full professorship. She had told them: "The exam is over, we have won, and I want you to know you are absolutely free; but, of course I would appreciate if you do it." According to the lawyer for the college, during the hearings, the plaintiff's request to the three students was clearly abusive and the Principal of the College mentioned one instance where she "attempted to exploit them (the students) for her private advantage" (p.6). In a previous situation, she had sent a registered letter to a student which said, "Since your evaluation is the only negative one out of the whole class, it is obvious that it is untrue and made deliberately in the intention of damaging my reputation, and perhaps, destroying my career. I therefore ask you to retract what you have written, and offer some sort of apologies. Should you decide not to comply with my request I would have to take some legal action" (p.8).[54]

As some legal scholars point out (Rebell, 1990), the significant aspect of faculty evaluation in general is that it serves a dual purpose:

> First, it is used to promote teacher training and development, while at the same time it serves to rate individuals for job termination. Such a system thereby sets up an inherent conflict of interest between the formative and summative functions, as the openness and cooperation necessary for staff development is in conflict with the self-protective, and adversary modes of rating and dismissal decisions. In the past, when summative decision making constituted only a minor part of evaluation activities, the underlying conflicts rarely came to the surface. However, since the adoption of educational reform legislation, evaluation techniques are increasingly being used to raise accountability standards by denying professional certification, retention, or promotion to those who do seem to meet acceptable standards. Thus, says Rebell, summative decisions, i.e., the type of decision which often leads to court cases, are becoming increasingly significant, and this increasing significance of summative evaluation decisions means more cases being brought before the courts (p.339-40).

As Rebell concludes, the adversary nature of the faculty evaluation process often contradicts the purpose for which they were initially developed: instructional effectiveness.

Release of SEF to Students and to The Public[55]

In addition to the administrative use of SEF, in recent years, other uses of SEF have become controversial, including releasing SEF to students and to the public.

> Summary: (76) Unlike most personnel records, SEF can be released to students and the public, on the grounds, that (77) students are not considered the general public, and (78) that SEF records are public and withholding them from public access does not outweigh the public interest in them.

While it is illegal to post a student's grades using a social security number or date of birth and a host of other confidentiality restrictions,[56] on a number of campuses, SEF data are openly published and sanctioned by some administrators and state government officials. In what many faculty see as an outrageous attempt to control the academic classroom, some state governments have sanctioned the release of SEF to the campus community, and in some cases to the general public, by publishing faculty student evaluations on the university's world wide web pages, thus making them not only available on campus but globally. One recent survey of accounting departments found that 11.4% of the respondents indicated that SEF scores are made available to students (Crumbley and Fliedner, 1995). Indeed, a search using "faculty evaluation" on the world wide web will return numerous examples of published SEF. All this while faculty are restricted from divulging information on students (see Pennsylvania State University, 1996). Articles are, however, beginning to appear that question the legality of publically releasing SEF (Robinson and Fink, 1996).

As I addressed in an earlier article (Haskell, 1997a), some faculty believe that due process and defamation issues are involved in SEF (see Crumbley, 1996), suggesting that faculty are entitled to at least the same rights as students. The Fourteenth Amendment, for example, requires due process before a public institution may deprive one of life, liberty, or property. A faculty member's reputation is considered a liberty right, and for tenured faculty the courts have pronounced the possession of tenure a property right. Presumably, any inappropriate action depriving faculty of these rights would be open to legal action. It has been suggested that if a university damages a faculty's reputation by publishing false and anecdotal data from SEF, faculty should be able to sue for libel or defamation. The concept of defamation typically refers to communication that causes a person to be shamed, ridiculed or held in contempt by others; to lower their status in the eyes of the community or to lose employment status or earnings or otherwise suffer a damaged reputation. Legally, while defamation is governed by state law, it is limited by the first amendment (Black, 1990).[57] According to one source, however, the courts have generally protected administrators from defamation charges resulting from performance evaluations (Zirkel, 1996). It would seem, however, that these older precedents applied when administrative evaluations were conducted in private and not publically distributed.

The release of personnel information is apparently allowed in no other phase of personnel or other key management functions. In typical personnel evaluations, professional validation studies are not permissible unless shown by professionally acceptable methods to be "predictive of or significantly correlated with important elements of work behavior which comprise or are relevant to the job or jobs for which candidates are being evaluated." In Title VII of the Civil Rights Act of 1964 the employer must meet "the burden of showing that any given requirement (or test) has a manifest relationship to the employment in question" (in *Griggs v. Duke Power Co.*, 401 U.S. 424 (1971). SEF data do not conform to these guidelines. It would seem that a university should be held responsible for insuring that data made public are valid. Given such apparent breaches of confidentiality and privacy, it will be instructive to see how the courts will continue to rule.

At the very least, SEF for administrative purposes and certainly the release of SEF data to students and the public create what the courts in a somewhat different context regarding academic freedom have termed a "chilling effect" on faculty behavior in the classroom (see *Keyishian v. Board of Regents*, 1967).

SEF in a Larger Context

In a larger context, it should be asked if there is reason to assume that what is happening in higher education is unique to the U.S. The answer is both "yes" and "no." It should also be asked if there is an additional (unintended?) effect of the administrative use of SEF, and the court's affirmation of its use, on the autonomy and future of the larger profession than those that have already been examined. The answer to this latter question is clearly "yes."

The Global Exporting of SEF

Given the largely peculiar U.S. institution of a generalized ideology of democracy and of consumerism, it might be suspected that SEF is an American institution. The fact is that like many aspects of U.S. culture, SEF is being exported around the world from Canada, to France, Germany, Great Britain, to Hong Kong, and beyond. Thus, an affirmative "yes" is the answer to the opening question of this section. The "no" answer is a more conditional one. Just as in the U.S., in other countries SEF appears to be linked to expanding enrollments to populations of atypical students, which often leads to a lowering of the quality of higher education.[58] For example in the late 1980s, policy makers in Hong Kong decided that in order to remain economically competitive, Hong Kong needed to develop its own research capabilities, train more professionals, and provide more opportunities in higher education to a much larger segment of the population. At that time only 3 per cent of high-school graduates went on to post secondary education; today its about 18 per cent. A recent government-commissioned study there concluded that the expansion of student enrollment in the university has been at the expense of academic quality (Hertling, 1996).

Similarly, in France, since its inception in the Napoleonic era, the *baccalauréat* has been aimed at the highest achieving students. The *baccalauréat* has earned a strong reputation in France and around the world, and has even inspired others to pattern their programs after it. Over the years, it has been criticized for being elitist. Before 1950, only 5 percent of an age cohort typically earned the *baccalauréat* in a given year. By 1992, 51 percent of the age cohort passed the *baccalauréat*.[59] Exams consist of both written and oral sections, with written tests taking up to four days with total testing time averaging up to 25 hours. The *baccalauréat* process has grown as a result of the government's desire to make it accessible to a larger, more diverse population. Questions are also being asked about the lowering of standards.

As two British researchers point out (Husbands and Fish, 1993), in some countries with politicians knowing that attacks on what is widely seen as the world of a privileged elite are electorally popular, there has been populist pressure encouraging them to intervene in the workings of higher education in order to reduce its apparent elitist nature. Enrollments in higher education have increased significantly in the UK, the Netherlands, France and Germany from a low former enrollment level by mandating new modes of entry such as flexible study programmes to encourage 'non conventional' students. In Britain, note Husbands and Fosh (1993),

> Even though student responses are considered an important aspect of the UK's approach to quality assessment of teaching, it is difficult to locate prescriptive statements that the gathering of such information should actually be by formal questionnaire. The questionnaire seems, almost surreptitiously to have become the most widely used means for the gathering of such views within UK universities--partly but (it is true) nor exclusively at the expense of other techniques. In addition, one does find references to, for example, liaison or consultative committees between students and staff, which may operate at institutional or departmental level. There are also institutions that favour gathering students' views on teaching using semi-directed or structured discussion groups lead by a facilitator who then prepares a report on the basis of comments given by participating students (e.g. Wisdom, 1991). Silver (1992). In his report on the present state of student feedback techniques in British higher education, makes a number of observations about general practices, as well as giving specific information about 14 institutions in England and Scotland that were examined in the course of his study. He describes the student questionnaire as "by far the most commonly used" method of

> obtaining feedback and offers a summary of arguments for and against such use, as well as discussing the variety of practices in its implementation (e.g. subject-specific versus comprehensive questionnaires or sampled responses versus total coverage) (p.100-101).

The authors go on to report that unsurprisingly, with the modest exceptions noted above, there is virtually no published research literature on issues of validity and bias in the use of SEF in German universities: "The extensive debates about the dimensionality of student assessment of teaching that have long raged among statistically oriented American, Canadian, and Australian researchers in higher education, and more recently in The Netherlands, seem as yet to have no complement in Germany" (p.103). As indicated in the opening to this paper, one of the few places there seems to be any question regrading SEF as an infringement on academic freedom is in Germany. Husbands and Frosh continue their view from Britain and Europe on SEF:

> It is a sad commentary on the gullibility of some people in the face of numerical data that it required the intervention of the courts to force the discontinuation of the more gross forms of this type of interpretation. As far as we know, there have been no comparable cases in European courts but, if European universities follow the American example of using student evaluations largely or exclusively for summative purposes, it is only a matter of time before there is external examination of the techniques being used, and of their suitability for the purposes for which they are intended. Certainly, if someone were, say, denied reappointment only or principally on the basis of the ratings that he/she had been given in students' assessments of his/her teaching, the institution concerned might expect to have the validity of its procedures on this subject examined extremely critically by the courts. Moreover, persistent denial of promotion merely on these same grounds alone might well lead to constructive dismissal proceedings instigated by the aggrieved individual. Again, the criteria and procedures would then come under very critical scrutiny (p.110).

Since Husbands and Fosh are largely addressing the non administrative use of SEF, they correctly conclude that they should not be interpreted as a calling for the complete discontinuation of SEF.

Legal and Administrative Default-Enforcement of Faculty Allegiance

Finally, and most importantly, there is an additional unrecognized contextual effect of the administrative use of SEF, and the courts' affirmation of its use, which in large measure undergirds the effects that have already been examined (e.g., SEF, grading, instructional methods, and standards). More specifically, this effect is the forcing faculty to shift from their traditional primary allegiance to the standards and norms of their discipline and to the larger profession of teaching to an allegiance with the standards and norms of the particular institution in which they teach. This is a major shift. The shift is the consequence of the SEF court rulings---and other issues addressed in this series of articles---which tend to attribute an inordinate degree of "good faith" in institutions and their administrators, giving them by force of law, the authority to decide (1) methods of instruction, (2) course content, and (3) grading practices.

The implication of SEF court rulings is this: that faculty are forced to teach to whatever level of student is enrolled in a particular institution or be subject to low evaluation by those students who may not have developed the ability to cope with traditional expected standards. The SEF is then used in tenure, promotion, and reappointment decisions. In this regard, recall the statement above by an administrator: "We're an open-admission university. A large fraction of the class was completely unable to compete" (Magner, 1995).

Unlike what many see as a voluntary compliance by faculty to pressures resulting from SEF,[60] the "unfriendly" court rulings presented in this series of articles legally enforces compliance to institutional acceptance of academic standards based on student evaluations. In my view, these rulings impinge on academic freedom by a quasi formal (i.e., by default) setting of academic standards. It should be noted that many of the rulings are not specific to cases involving SEF as similar rulings have been extant in other educational contexts for some time. When applied to SEF, however, such rulings seem to widen the context of their original application.

This legally enforced shift in faculty allegiance not only affects instructional matters, but arguably changes the nature of the faculty employment relationship to the institution. Traditionally, faculty have not considered themselves employees of the institution in the standard sense of an employee in a business corporation; the relationship has been generally seen as an independent contractor. There are crucial differences, however, that render faculty not quite like traditional employees in business. The academic governance systems of universities render faculty "partners" in the management of the institution, though recently there have been suggestions to reduce the traditional role of faculty governance.[61]

In terms of contract law, however, faculty are employees of the institution, and as I noted above, legally, faculty set academic standards only in so far as they are accepted by specific contractual agreement with the institution (see section: Who Is the University That Sets Academic Standards?). The situation is not quite as clear-cut as it may appear. It is unclear, for example, if courts would accede to a formally stated set of faculty standards and norms if not accepted as a contractual agreement by the institution. As Kaplin suggests, courts sometimes could and indeed might "accede to professional standards and norms of faculty even though such standards and norms are not embodied in a formal contract. Apparently, courts could so rule under the theory that such standards and norms are part of local or national 'academic custom and usage' " (personal communication, August 11, 1997. See also Kaplin and Lee, 1995, section. 1.3.2.3). But the courts apparently seldom accede to such informal norms. In any event, at the very least, the legal rulings seem to shift the status of faculty---in effect, if not in total fact---from an independent professional to an employee. If so, then additional changes in traditional faculty prerogatives will likely follow, further eroding academic freedom.

As already noted, higher education is increasingly being seen as a *business*. The university considered as a business carries with it the attendant and associated ideas of students as *consumers* in an *educational marketplace* and faculty as *employees* of the business. Thus, the enforced shift in faculty allegiance can lead to a domino effect in eroding academic freedom and tenure. It is a shift with crucial implications of its own which needs to be addressed by higher education.[62]

Once again, as I discussed in the above section, "Who Is the University That Sets Academic Standards?" if faculty are to regain and maintain their academic role they need to formally and in detail address SEF, teaching methods, and grading issues in their contractual agreements and handbooks. As Kaplin and Lee (1995) have suggested relative to academic freedom in general, "It is especially crucial for institutions to develop their own guidelines on academic freedom and to have internal systems for protecting academic freedom in accordance with institutional policy" (Section 3.6.1, p. 192). This would appear to be especially true for SEF, teaching methods and grading policies.

Conclusion

The issue of what sustains SEF and its associated problems of academic standards and of maintaining student tuition is a complex one, including positive reinforcement patterns by all parties involved. Assuming that the situation calls for at least some modicum of change, it can not be accomplished on an individual level; it has to be accessed on a systems level. On a macro level, this means changing cultural values about education, the university's economic orientation, administrative practices, student orientation to learning, and faculty collective action. We must change the reward structures so that each party does not gain from the situation. Currently, parents gain when their children who might not otherwise earn a college degree do acquire one; college presidents gain by demonstrating to boards of trustees that they are constructing new buildings; trustees gain by demonstrating an economically viable institution, other administrators gain because they can show the president growth within their own administrative units; other units within the university like academic departments gain because department budgets tend to be based on student enrollment numbers; students gain because they do not have to study hard to attain an A or B grade-point average;[63] and finally, faculty gain because they are rewarded both by student evaluations, and administrators. This is what is called a closed, mutually rewarding, escalating system with little to no restraining feedback. Systems engineers would recognize this as a run-a-way system. Unfortunately, change may have to come from external sources like accrediting agencies.

Notes

1. Address correspondence to: Robert E. Haskell, Ph.D., Professor of Psychology, Department of Social and Behavioral Sciences, University of New England, Biddeford, ME 04005. Email rhaskell@javanet.com. I would like to thank Professor John Damron, of Douglas College for continually providing me with sources, support, and advice, and especially Professor William A. Kaplin, School of Law, Catholic University of America, for his invaluable legal counsel and for reading a draft of this paper. The conclusions reached in this paper do not, of course, necessarily reflect the views of those who contributed to its development.

[BACK to document]

2. To question SEF often meets with emotional reactions. In a response to my first article, Michael Theall (1997), Director of the Center for Teaching and Learning at the University of Illinois at Springfield, opens his critique of my article by lamentably describing my piece as rhetorical, characterizing it as an example of

> (1) faculty who "fulminate" against SEF, (2) as "simplistic," (3) "loaded with misinterpretations of the literature," (4) as "mythology," (5) exhibiting an "ignorance of evaluation/measurement literature," as containing (6) "sweeping generalizations," (7) "misinformation," (8) as "simply ridiculous!" (9) "ripe" with "hysterical rhetoric," (10) as assuming a "mythical group of better students" of some bygone era, asserting that (11) SEF "are the cause of grade inflation," as suggesting (12) "we do away with ratings." Continuing, Theall wrote that (13) "Perhaps the weakest part of his article is what isn't there: constructive suggestions for improvement," that I (14) "suggested that ratings are a violation of academic freedom," and finally---but not exhaustively---(15) that "Academic freedom has been defined in many ways, but never before in a way that suggests the construct (tradition? principle? tenet?) is vulnerable to the influence of student ratings."

In response (Haskell, 1997d), I suggested that the author at least got the last two items correct. I agree with Hamilton's (1997) observation that "One of the greatest contributions an academic can make is an honorable defense of the principles on which the university rests" (p.19).

[BACK to document]

3. To further demonstrate the negative reaction to question SEF, a more recent negative response to my article is Marsh and Roche (1997). In grossly misrepresenting my thesis (to which I had no opportunity to respond) they said:

> **Experimental field studies.**
> Marsh (1984, 1987; Marsh & Dunking, 1992; also see Abram I, Dickens, Perry, & Leventhal, 1980; Howard & Maxwell, 1982) reviewed experimental field studies purporting to demonstrate a grading-leniency effect on sets but concluded that the research was weak and flawed. In marked contrast, *Haskell (1997) summarized work implying that these studies provide good evidence for a grading leniency effect, even suggesting an implicit collusion among SET researchers to hide this conclusion.* It is important to counter such dubious but *popular interpretations*, because the use of deception in these studies would presumably fail to meet current ethical standards, making the studies difficult to replicate or refine. Here, we briefly elaborate four crippling weaknesses of these studies by Chacko (1983), Holmes (1972), Powell (1977), Vasta and Sarmiento (1979), and Worthington and Wong (1979), and one subsequent study by Blunt (1991). p.1191 [italics added]

In fact, the only place I cite Marsh, etc., is in a footnote (# 3, see below). First Marsh and Roche misread the footnote suggesting that I cite Marsh (1984, 1987; Marsh & Dunking, 1992; also see Abrami, Dickens, Perry, & Leventhal, 1980; Howard & Maxwell, 1982) relative to grading leniency. My footnote clearly refers to validity of SEF, not grading leniency. Later in the paragraph, I refer to Greenwald's work on grading leniency:

> 3. Since the issue of SEF validity, in terms of learning, is so central a few observations of the literature are necessary. Greenwald and Gillmore (1996) have categorized some of the significant reviews and empirical research that find in favor of validity of SEF as measures of quality of instruction, for example, Cashin (1995), Cohen (1981), Feldman (in press), Howard, Conway, and Maxwell (1985), Howard and Maxwell (1980, 1982), Marsh (1980, 1982, 1984), Marsh and Dunking (1992), and McKeachie (1979). Reviews and empirical critiques that are critical of the validity of SEF include, Chacko (1983), Dowell and Neal (1982), Holmes (1972), Powell (1977), Snyder and Clair (1976), Vanta and Sarmiento (1979), and Worthington and Wong (1979). Positions, suggesting cautious support for validity of SEF while at the same time expressing concerns about the adequacy of their support, include, Abrami, Dickens, Perry, & Leventhal (1980). The recent methodologically sophisticated research of Greenwald (1996), and Greenwald and Gillmore (1996) find strong evidence inconsistent with the common dismissive interpretation of the relationship between SEF and high student grades as reflecting a relationship between amount learned and student ratings. (Haskell, 1997).

Moreover, Marsh and Roche claim that I implied "an implicit collusion among SEF researchers." Nowhere, however, in my paper do I suggest collusion among researchers. In addition they refer to my piece as "popular." What does this mean? If popular means non statistical, it's a strange definition. If it refers to a piece that emphases the implications and contexts of an issue---a legitimate and important approach to an issue---again, a rather strange definition of "popular." My articles are, after all clearly "policy" pieces. Finally, what Marsh and Roche refer to was in a footnote and far from the main thesis of the paper. I consider Marsh and Roche's comment on my article not just a "misinterpretation" or a "mis reading," but a clear misrepresentation. I also consider it unscholarly in that they do not document their misrepresentation (Perhaps Marsh and Roche were obliquely reacting to my response to Theall's commentary (see previous endnote above). And unlike Theall, I had no opportunity to reply and defend my integrity.

[BACK to document]

4. For a recent dialogue on SEF in *Change: The Magazine of Higher Learning*, see Williams and Ceci (1997), Trout, (1997). See also Winston (1997) for an economic analysis of long term institutional costs of lowering academic admission standards. For a kinder and gentler view of SEF see Gold (1997).

[BACK to document]

5. For convenience of exposition, unless otherwise specified, I will use the simpler "SEF used administratively" for the more cumbersome "SEF as used administratively in reappointment, tenure, and promotion decisions."

[BACK to document]

6. Neil Hamilton (1977), Trustee Professor of Regulatory Policy at William Mitchell College of Law, has observed that "Many academics do not seem aware of our continuing struggles for the principles of academic freedom and tenure, nor can they make a reasoned defense of the principles" (p.16).

[BACK to document]

7. My thanks to Michiel Horn, Professor of History, Glendon College of York University, Toronto, Canada, for his collegiality in calling my attention to Knapper's chapter. In addition, while researching the first in this series of articles (Haskell, 1997a, and as I indicated in endnote # 9 of that paper), according to Husbands and Frosh (1993) at the time of their writing, there was a lively debate in German higher education over whether student evaluation of teaching is an invasion of academic

autonomy. It is perhaps no coincidence that academic freedom is considered by many scholars to have its origins in Germany.
[BACK to document]

8. AAUP Committee C on College and University Teaching, Research, and Publication. It was adopted by the Council of the American Association of University Professors in June, 1975, and endorsed by the Sixty-first Annual Meeting as Association policy. They also state: "An important and often overlooked element of evaluating teaching is an accurate description of a professor's teaching. Such a description should include the number and level and kinds of classes taught, the numbers of students, and out-of-class activities related to teaching. *Such data should be very carefully considered both to guard against drawing unwarranted conclusions and to increase the possibilities of fairly comparing workloads and kinds of teaching*, of clarifying expectations, and of identifying particulars of minimum and maximum performance. Other useful information might include evidence of the ability of a teacher to shape new courses, to reach different levels and kinds of students, to develop effective teaching strategies, and to contribute to the effectiveness of the individual's and the institution's instruction in other ways than in the classroom" (Itlaics added). As the data in these articles demonstrate, these AAUP guidelines are seldom adhered to.
[BACK to document]

9. See also Kaplin and Lee (1995) for a historical overview. "3.7.1. General concepts and principles. The concept of academic freedom eludes precise definition. It draws meaning from both the world of education and the world of law. Educators usually use the term 'academic freedom' in reference to the custom and practice, and the ideal, by which faculties may best flourish in their work as teachers and researchers (see, for example, the AAUP's 1940 'Statement of Principles on Academic Freedom and Tenure,' in AAUP Policy Documents and Reports (AAUP, 1990), 3-10). Lawyers and judges, in comparison, often use 'academic freedom' as a catch-all term to describe the legal rights and responsibilities of the teaching profession, and courts hearing such cases attempt to reconcile basic constitutional principles with prevailing views of academic freedom's social and intellectual role in American life. Moreover, academic freedom refers not only to the prerogatives of faculty members and students but also to the prerogatives of institutions ('institutional academic freedom' or 'institutional autonomy'). "See also Menand (1996) for interesting discussion of the early history of academic freedom in relation to academic disciplines.
[BACK to document]

10. As in my previous papers on SEF, to render them manageable I have set a number of definitional and constraining parameters. Accordingly, (1) the term "court" as used here includes rulings by legally constituted Arbitration Boards; (2) since this paper is only concerned with how these legal bodies have ruled on various aspects SEF data, I do not distinguish between state courts, federal courts, or arbitration boards; (3) neither does this paper deal with the multiple legal variables that define and distinguish a legal action, influence, or outcome in a particular case, such as the particular statute or other source of law being applied, e.g., the cause of action being asserted, the prescribed prima facie case, the allocation of burdens of proof, and the standards of judicial review; (4) nor will the paper be concerned with the complex legal reasoning on which the rulings and outcomes were based; (5) finally, I address the legal material not as a legal scholar but from the "reasonable man" standard; (6) the purpose of which is to inform future educational policy and legal change.
[BACK to document]

11. In *Lovelace v. Southeastern Massachusetts University* (1986), the courts ruled: (1) "It is important to note what plaintiff's first amendment claim is and to separate speech from action. Plaintiff has not contended that he was retaliated against simply because he advocated that the university elevate its standards.... Plaintiff's complaint instead is that he was retaliated against when he refused to change his standards" (p.425); (2) citing other cases, the court rejected his contention that a university teacher has a first amendment right to disregard established curriculum content, that

the first amendment does not prevent a university from terminating an untenured faculty whose pedagogical style and philosophy does not conform to those of the school's administration; (3) that pedagogical style and philosophy "is a policy decision which, we think, universities must be allowed to set" (p.426). Further, the court ruled that (4) "We will assume for purposes of this opinion that plaintiffs refusal to lower his standards was a substantial motivating factor (see *Mount Health Board of Education v. Doyle*, 429 U.S. 274, 283-284, 97 S.Ct. 568, 574-575, 50 LEd.2d 471 (1977) in the decision not to renew his contract.

In *Carley v. Arizona Board of Regents* (1987), (4) Carley claimed as "protected speech" his teaching methods where his goal in his commercial art course was to promote a business atmosphere by requiring attendance, promptness, and self-reliance, and he required them to meet deadlines. The court ruled (19) his teaching style is not a form of speech protected under the First Amendment. (20) Decision not to retain a non tenured instructor, even if based, in part, upon student evaluations expressing disapproval of his teaching methods, did not violate instructor's First Amendment right to academic freedom; (21) Carley was not denied a contract because of expressing unpopular opinions or otherwise presenting controversial ideas to his students. (22) Thus, we conclude that the decision not to retain Carley, even if based, in part, upon student evaluations expressing disapproval of his teaching methods, did not violate his first amendment rights (p.1103).

[BACK to document]

12. Courts often cite the *Regents of the University of California v. Bakke*, 438 U.S. 265, 312, 98 S.Ct. 2733, 2759-2760, 57 L.Ed.Sd 750 (1978) (the "four essential freedoms" of a university are "to determine for itself on academic grounds who may teach, what may be taught, how it shall be taught, and who may be admitted to study."

[BACK to document]

13. One of the major variables in shaping legal rulings on faculty matters is whether a faculty member's constitutional rights have been abridged. A further variable is whether faculty teach at a private or state college. Because a state operated college is governmental First Amendment Rights apply, whereas at private institutions faculty do not have the same degree of protection.

[BACK to document]

14. The court referred to: *Millikan v. Board of Directors of Everett School District No. 2*, 1980; *Adams v. Campbell County School District*, 1975; *Riggin v. Board of Trustees of Ball State University*, 1986.

[BACK to document]

15. This course is sometimes labeled small group processes, group communication, interpersonal interaction, or T-group. I have been teaching such a course for years. Most students tend to like the course, because I teach it in a relatively nondirective, experiential manner, requiring two papers analyzing the group dynamics, not tests. Consequently I receive my highest student evaluations in this course. I once taught this course to students in a nursing program. Preferring more structure than many students, they did not like the course. I was not asked to teach the course again. It was taught thereafter in a lecture format in the nursing department. This is most unfortunate for their future patients, as the course is a "practice" course in interpersonal skills. During the late 1960s and early 1970s, experiential group courses (some then called "encounter" or "sensitivity" groups) were objected to by some religious groups because some of the courses are directed at affective, not just intellectual, change. This brings up the additional question of academic freedom involving instructional methods of faculty who teach courses that service the curricula of other departments. Who has the legitimate control of content and teaching methods in service courses?

[BACK to document]

16. In arguing faculty rights, I am not suggesting unlimited rights. Just as in the legal concept of free speech, there are certain narrowly defined limits, so too in the faculty right choose a teaching method there are limits. A recent issue of the *Chronicle of Higher Education* (1997, November 14)

reports the case of a faculty who was dismissed because of his teaching method. At least prima facially as described, this could be an example of the limits of faculty right to use whatever teaching method s/he deems appropriate.
[BACK to document]

17. For a more extended critique and criticism of Parate, Kaplin and Lee suggest: D. Sacken, "Making No Sense of Academic Freedom: Parate v. Isibor," 56 West's Educ. Law Rptr. I IO7 (Jan. 4, 199O).
[BACK to document]

18. For other nuances of court rulings on grading, see *Susan M v. New York Law School*, 544 N.Y.S.2d 829 (N.Y. App. Div. 1989). A law student dismissed for inadequate academic performance sought judicial review of her grades in her constitutional law and corporations courses. The court said:

> At least when a student's very right to remain in school depends on it, we think the school owes the student some manner of safeguard against the possibility of arbitrary or capricious error in grading, and that, in the absence of any such safeguards, concrete allegations of flagrant misapprehension on the part of the grader entitle the student to a measure of relief [544 N.Y.S.2d at 831-32].

The court then outlined the kind of review it believed to be appropriate. Had it been upheld on appeal, the outline would have subjected faculty reasoning processes in grading to judicial scrutiny. It said:

> At issue is not what grade petitioner should have received but whether the grade received was arbitrary and capricious; not whether petitioner deserved a C+ instead of a D in Corporations but whether she deserved a zero on this particular essay; not the quality of petitioner's answer but the rationality of the professor's grading [544 N.Y.S.2d at 832].

The court returned the issue to the law school for further consideration, asking the school to provide reasonable assurance that the zero assigned on the student's essay was a "rational exercise of discretion by the grader" (544 N.Y.S.2d at 832). The school appealed, and the state Supreme court unanimously reversed the appellate court, reinstating the trial court decision. The court strongly supported the academic deference argument made by the school, stating:

> Because [the plaintiff's] allegations are directed at the pedagogical evaluation of her test grades, a determination best left to educators rather than the courts, we conclude that her petition does not state a judicially cognizable claim [556 N.E.2d at 1105].

After reviewing the outcomes in earlier challenges, the state's highest court stated:

> As a general rule, judicial review of grading disputes would inappropriately involve the courts in the very core of academic and educational decision making. Moreover, to so involve the courts in assessing the propriety of particular grades would promote litigation by countless unsuccessful students and thus undermine the credibility of the academic determinations of educational institutions. We conclude, therefore, that, in the absence of demonstrated bad faith, arbitrariness, capriciousness, irrationality or a constitutional or statutory violation, a student's challenge to a particular grade or other academic determination relating to a genuine substantive evaluation of the student's academic capabilities, is beyond the scope of judicial review [556 N.E.2d at 1107].

The court concluded that the claims concerned substantive evaluation of academic performance, and therefore refused to review them.
According to Kaplin and Lee (1995, see section 4.7.1. Awarding of Grades and Degrees 473), students have not typically prevailed in challenging grades. They suggest the following summary of legal challenges to academic judgments and a review of the *Susan M* case, see Note, "Student Challenges to Grades and Academic Dismissals: Are They Losing Battles?" 18 *J. Coll. do Univ. Law*

577 (1992). See also F. Faulkner, "Judicial Deference to University Decisions Not to Grant Degrees, Certificates, and Credit—The Fiduciary Alternative," 40 *Syracuse L. Rev*. 837 (1990), and T.A. Schweitzer, " 'Academic Challenge' Cases: Should Judicial Review Extend to Academic Evaluations of Students?" 41 *American U. L. Rev*. 267 (1992). *Susan M* is also humorously reviewed in verse by R.E. Rains in 40 1. *Legal Educ*. 485 (1990) and 43 *J. Legal Educ*. 149 (1993).
[BACK to document]

19. Once again, when discrimination charges are a part of grading complaints, the courts engage in closer scrutiny of grading practices. See (Haskell, 1997c) the section on Disparate Treatment and Disparate Impact.
[BACK to document]

20. In *William Sypher v. Vermont State Colleges Faculty Federation* (1982), it was observed (9) with regard to the "political" aspect of the case that Sypher had written a letter in defense of his student rating level at the college. In it, he said "It is certainly distressing when very good is not good enough, especially at a college with a modestly-talented student body that often discourages efforts at subtlety, wit and deeper penetration of subjects." (10) The Board responded to this letter saying, "Other actions and statements by Grievant constituted legitimate reasons for not retaining him. In a May, 1980, letter to Dean Beston, Grievant expressed his contempt for Castleton students" (p.135), (11) concluding, "Accordingly, we find credible the College's contention that Grievant was not reappointed because of his teaching effectiveness [Italics added] (p.135).V.
In *Lovelace v. Southeastern Massachusetts University* (1986), the court said, (4) whether "a school sets itself up to attract and serve only the best and the brightest students or whether it instead gears its standard to a broader, more average population is a policy decision which, we think universities must be allowed to set....matters such as course content, homework load, and grading policy are core university concerns" (p.424).
In *Robert Kramer v. The President of the University of British Columbia* (1992), the Board said (30) "One perceptive student noted that some of the unhappiness came from the fact that the levels of Japanese language ability were badly divided; some found it easy, others very hard" (p.12).
[BACK to document]

21. See also Kaplin and Lee (1995), sections. 2.1, 2.2.1, 2.2.2, and 2.2.3.
[BACK to document]

22. Interesting in this regard, recently (Strosnider, 1996) two Pace University students sued the university on the grounds that because the course was too difficult they had to drop it, and therefore they should not have to pay for it. The judge who initially heard the suit found in favor of the students, awarding each of them $2,065.31 for the tuition they had paid, plus damages. An appellate court overturned the decision, ruling that the lower-court judge had "improperly engaged in judicial evaluation of a course of instruction that the courts of this state have consistently held is the proper domain of educators and educational institutions entrusted to the task." What is interesting in terms of this paper is that the Pace university spokesperson was quoted as saying the importance of the case was that it "*had to do with the ability of faculty members to run a class in the manner they feel appropriate*" [italics added].
[BACK to document]

23. In one classic and contemporary academic freedom statement (Morrow, 1968), it was noted that "legal control is usually vested in a non academic board, council, or court, distinct from the faculties of the institutions. Although their control is legally unlimited, these boards are subject to the restraint of public opinion, of academic custom and precedent, and of accrediting bodies and professional organizations" (p.6) Morrow also notes,

> Academic freedom, in its primary sense, is the freedom claimed by a college or university professor to write or speak the truth as he sees it, without fear of dismissal by his academic superiors or by authorities outside his college or university. In a secondary sense, the term denotes the corporate freedom claimed by an

> institution of higher learning to determine its policies and practices, without restraint from outside agencies. This latter usage is clearly distinct and derivative; for such corporate autonomy derives its justification ultimately from the services performed by the scholars whose activity it exists to foster and protect, while, on the other hand, the freedom of the individual scholar often requires protection from the pressures of his own institution, as well as from outside forces (p.4).

[BACK to document]

24. I was unable to determine from national unions what percentage of contracts preclude the administrative use of SEF. Perhaps others will be more successful in locating this information. It is important information because the false belief by many administrators and faculty that nearly all schools use SEF administratively is part of what maintains the practice and inhibits change. As I indicated in Haskell (1997c, endnote # 17), some schools (a) do not mandate SEF, (b) do not require its administrative use, or (c) constrain its use in various ways. For example, at the University of Guam (Blum D. E.,1990, October 3), an arbitration board ruled that SEF should not be used unless (1) students are made aware of the purpose and ramifications of their evaluations, and (2) that evaluations should not be anonymous. A Rider University, the faculty agreement stated "The College may not use course evaluations for purposes of discipline, promotion, or tenure, unless introduced for such purposes by the faculty member." At Western Michigan University, the faculty agreement stated "Only the ratings shall be included in all promotion, reappointment, merit, and tenure recommendations, together with such other evaluations of teaching competence as may be employed by faculty members and made available. Western agrees to consider all the evidence of teaching competence that is presented in evaluating teaching faculty and shall not use unsubstantiated structured comments in personnel decisions." I have also been informed by a colleague at St. John's University (New York) that, while SEF is required, it is not used administratively (though this may be changing). I suspect there are many more schools (likely those who have union contracts) that do not use SEF administratively or who limit its use.

[BACK to document]

25. I would like to thank AAUP attorney Patrick B. Shaw for referring me to Ms. Linda Lott, Administrative Coordinator, Hofstra Univeristy Chapter, AAUP, who conducted a search of the new collective bargaining contract database (March 21, 1997). Searching the database was conducted with "several key words that relate to academic freedom, teaching methodology and student evaluations. Virtually no contract was found that specifically dealt with faculty right to set course content and teaching methods. I also would like to thank Ms. Maureen Webb, attorney for the Canadian Association of University Teachers (CAUT) for her assistance.

[BACK to document]

26. Cited in *Grievance of William Sypher and the Vermont State Colleges Faculty Federation*, 5 VLRB 102 (1982) (p.125). Knapper (1977) notes an early

> attempt to provide safeguards for the use of evaluation data was made by a committee on evaluation established by the University of Ottawa, whose report is well worth reading (University of Ottawa, 1974). The committee recommendations allow for confidentiality of data, and consultation with the faculty member on the criteria to be used in assessing his performance. Faculty must be given due notice of any unfavourable assessments, and an opportunity to comment and demonstrate improvement before any information is passed on to their administrative supervisor. They must also be kept informed of evaluation information placed in their file, and student comments will only be placed in the file if the students have taken up complaints directly with the instructor first. Finally, there is provision for an appeal procedure (p.202).

[BACK to document]

27. The 1973 AAUP report on faculty tenure notes, "The faculty of the institution...must be the

source for the definition and classification of standards of professional conduct and must take the lead in ensuring that these standards are enforced....[and] accept their full corporate responsibility for the integrity of the profession" http://www.aaup.org.

[BACK to document]

28. John Dewey, (1976, orig 1902) wrote in this regard,

> Implicit, if not explicit, obligations are assumed. In this situation, conflict between the two concerns of the university may arise; and in the confusion of this conflict it is difficult to determine just which way the instructor is morally bound to face. Upon the whole it is clear however....We can insist upon one hand that the individual must be loyal to truth, and that he must have the courage of his convictions; that he must not permit their presumed unpopularity, the possibly unfavorable reaction of their free expression upon his own career, to swerve him from his singleness of devotion to truth (p.54).

[BACK to document]

29. In *Robert Kramer v. The President of the University of British Columbia* (1992), it was noted that (2) The department head viewed Kramer's 1989-90 course evaluations "with some alarm" and that a number of students had stated that Dr. Kramer's teaching would cause them to stay away from the Asian Studies department, and (6) since the course was the general introduction to the subject and the Department, such negative comments were of great concern to the department head.

[BACK to document]

30. The Harvards, Dukes, Yales, Stanfords, and popular media---as indicated in one popular book entitled *Profscam* (Sykes, 1988)---notwithstanding, most faculty, as numerous national surveys document, do not spend the majority of their time on research as apposed to teaching. It seems to be a failure of proportional reasoning and judgement that on many campuses and at many conferences on higher education that research is increasingly devalued by calling fourth as the justificatory example faculty at mega research universities who spend reduced time teaching, thereby suggesting that most faculty in the teaching trenches are not primarily concerned with classroom teaching. In addition, faculty and administrators at small colleges often justify, and indeed discourage research by pointing to the literature on the "over emphasis" on research using mega research universities as the examples, as if the same logic applied to them.

[BACK to document]

31. It is telling, I think, that the court should take objection to this faculty member's assessment of the level of student ability in his classes.

[BACK to document]

32. I am not implying that administration is the "bad guy" and faculty is the "good guy." Reality is not that simple. As I have stated previously, "This is not intended as a blanket apologia for academia. There are many problems within the academy. In many other areas, I am a severe critic of my colleagues collective behavior" (Haskell, 1997a, endnote 4). In fact, faculty are in considerable measure collectively responsible for the current state of affairs regarding grade inflation and lowered standards. It is no secret that, historically, faculty on college campuses have tended to be disorganized, and that this disorganization has contributed to loss of control over standards (see for example Grose, 1997; Power, 1997; Shattuck, 1997). In a similar context, Scriven (1996) in his provocatively entitled article, "The Treason of the Intellectuals," has lamented, "Disciplined inquiry should begin at home, and it is a wretched commentary on academic intellectuals that they have been unwilling to apply it to the very activity that earns them a living and made them capable of earning that living." Faculty must do more than complain. Given this, administrators are subjected to a different reward system than are faculty. They are rewarded for enrollments and constructing new facilities, with the latter requiring the former.

[BACK to document]

33. Indicative of the interpretation of SEF remarks and their administrative use, on one recent professional Internet discussion group, a faculty member offered his particular experience, noting that student comments can be used for any purpose. One student commented he was so brilliant that he should quit teaching at the undergraduate level and teach graduate school. The administration interpreted the student's comment to mean that the faculty was not doing a good job, that the comment meant that he was talking over their heads (Due to privacy issues I will not cite the Listserv. I will privately provide the URL upon request).

[BACK to document]

34. More strictly speaking, in discrimination cases, courts apparently do begin by trusting universities. They set aside this trust only after a plaintiff has succeeded in proving a *prima facie case* of discrimination. Similarly, if a plaintiff can make a *prima facie* case regarding, say, a free speech violation or a due process violation in a pedagogy or academic standards case, the courts would also apparently set aside their trust. Once more I am indebted to Bill Kaplin (personal communication, October, 6th, 1997) for his legal expertise. While professor Kaplin is undoubtedly correct, in my view with respect to the specific SEF cases reviewed in this series of articles, it could reasonably be concluded that the courts require more evidence to establish a *prima facie* case than with discrimination cases.

[BACK to document]

35. In a recent and ongoing denial of tenure on the basis of SEF, an attorney reviewing the findings summarized the case by saying "that in this case the student evaluations were almost uniformly unfavorable and were relied upon heavily by the University to justify the denial of tenure (*The University took the view that students were customers of the University*). Based on the student evaluations and the high failure rates in the professor's courses, the University alleged that he was a poor teacher because he pitched his engineering courses at too high an academic level, put too great an emphasis on mathematics, and taught at too fast a pace. *Our argument was that it was within the professor's academic freedom to choose the content, emphasis and pace in a course within the parameters of international and national standards [which were provided by expert testimony]*, the course outline approved by Senate and the prerequisites for the course. To support the argument we pointed to the Faculty Handbook, which gave a broad guarantee of academic freedom" (italics added). One campus advertises with the slogan "Your Success Is Our Business." See http://www.tulsa.cc.ok.us/register.html#Procedures.

[BACK to document]

36. One court judge noted (*EEOC V. Franklin & Marshall College*, 1985) "I do not agree with the majority's assumption that academic institutions are the same as any other employer. At least insofar as their administrative and governance structures are concerned, colleges and universities differ significantly from garden variety private employers. In the context of application of the provisions of the National Labor Relations Act the Supreme Court has counseled that principles developed for use in the industrial setting cannot be 'imposed blindly on the academic world.'" *NLRB v. Yeshiva University*, 444 U.S. 672, 681, 100 S.Ct. 856, 861, 63 L.Ed.2d 115 (1980) p.120.

[BACK to document]

37. McMurtry delineates the inappropriate use of the consumer metaphor in education.

> The best product on the market, as we know, is the one which is the most 'problem-free' for its purchaser--delivered ready made for instant easy use','guaranteed replacement' if it does not work, and 'repaired cost-free' whenever it needs maintenance attention. The best education, on the other hand, is the opposite on all standards of excellence. It cannot be produced or delivered by another at all, is never ready-made nor instant, and cannot be guaranteed replacement or service cost-free if it is not working. The higher the standards it has, the less it can be immediate in yield, the more work it demands of its owner, and the more its failures must be overcome by its possessor's own work. An education can never be 'problem free', and poses ever deeper and wider problems the higher the level of excellence it achieves. Freedom in the market is the enjoyment of whatever one is able to buy from others with no questions asked, and profit

> from whatever one is able to sell to others with no requirement to answer to anyone else. Freedom in the place of education, on the other hand, is precisely the freedom to question, and to seek answers, whether it offends people's self-gratification or not....What is the best policy for buying a product--to assert the customer's claim 'as always right'--is the worst possible policy for a learner. What is the best policy for selling a product--to offend no-one and no vested interest--maybe the worst possible policy for an educator. The principles of freedom here are contradictory, and become the more so the more each is realized (p.213-214).

[BACK to document]

38. In Morrow's (1968) classic statement, he says,

> An ordinary citizen who expresses unpopular opinions may lose customers if he is a merchant, clients if he is a lawyer, patients if he is a physician, advertisers or subscribers if he is the editor of a newspaper, or suffer other forms of social or economic penalty resulting from disapproval of his expressed opinions. The university professor, in some degree, suffers similar consequences; but where academic freedom is recognized, he is protected from the gravest of them, namely, the loss of his position. The justification of academic freedom must therefore be sought in the peculiar character and function of the university scholar (p.6).

[BACK to document]

39. In this regard, Dewey recognized very early on (1976, orig. 1902):

> A new type of college administration has been called into being by the great expansion on the material side. A ponderous machinery has come into existence for carrying on the multiplicity of business and quasi-business matters without which the modern university would come to a standstill. This machinery tends to come between the individual and the region of moral aims in which he should assert himself....Now the need for money is not in itself external to genuine university concerns; much less antagonistic to them. The university must expand in order to be true to itself, and to expand it must have money. The danger is that means absorb attention and thus possess the value that attaches alone to the ultimate educational end. The public mind gives an importance to the money side of educational institutions which is insensibly modifying the standard of judgment both within and without the college walls (p.62-63).

[BACK to document]

40. It is unfortunate with regard to this issue that both ideology and politics have reached the level where views are automatically labeled "liberal" or "conservative." Since the publication of my first SEF article, I have been called an "arch conservative" by some of my colleagues. I have always thought of myself as a "liberal," though on academic matters I am perhaps more conservative than liberal. (See also Peter Sacks' book for his response to being similarly labeled). Not only are these labels too simple, but also such labels preclude addressing the issues involved.

[BACK to document]

41. Given the current political climate and some responses to this series of articles (see endnotes # 2 & 3), it is perhaps necessary to clearly state that the discussions around grade inflation and lowered admissions standards are not meant as "code words" for blaming minorities as one Harvard professor has explicitly voiced (Mansfield, 1993). The argument is that professors tend to "overgrade" all students in order to justify and obscure their "overgrading" of black students. Even on the general face of it, this does not seem a valid cause of any grade inflation or lowered admissions standards. The argument assumes that the academic level of black students is lower (*a la* the "Bell Curve" or lack of academic preparation?) than that of white students. *For the sake of this argument only*, let's assume that blacks *on a group average* do show lower ability than whites. I don't think this would make one iota of difference in the grade inflation problem. This is why: First, faculty do not just have to obscure the overgrading of blacks. The wholesale lowering of admission standards across

the country forces overgrading to be an equal opportunity system---and mostly for whites. Blacks are simply swooped up in the more general admissions and grading inflationary situation that admits very low level white students. Second (and again even for the moment assuming the average lower level of black students) we know that---just as with whites---many of the black students are of high quality, which would argue against their assumed contribution to grade inflation. It is interesting that the Harvard professor did not say that grades were being inflated because of giving blacks higher grades, but because the lower academic level of black students are forcing faculties to grade everyone higher in order to hide the "fact" that professors are forced to pass black students. This is a slightly more sophisticated argument than the typical one that simply states that grades are being inflated because of having to give passing grades to blacks. But many people will probably not catch this distinction. The Harvard professor apparently knows that the simple argument that grades are being inflated simply because of having to give passing grades to blacks will not past musters because, third, given the small number of blacks in higher education, any assumed contribution by blacks to grade inflation and lowering of standards, renders any consequent rise in grades due to black admissions an imperceptible blip on the inflated grading curve. Finally, to blame blacks for grade inflation, doesn't address the more obvious fact of grade inflation on campuses without any black students.

[BACK to document]

42. The University of Washington was reportedly chided by an accrediting agency because 70 per cent of the grades were As or Bs, up from 60 per cent in 1983 (Shea,1994). On many campuses the average grade is now an A- or B+. Evidence of grade inflation is indicated by the population of students since 1987 with A plus, A, and A minus grade-point averages having increased from 28% to 37%, while at the same time SAT scores have decreased an average of 13 points on verbal (but only 1 point on math). The average for all SAT takers is 3.22 on a four-point scale, a considerable increase above an average of 3.07 in 1987. The continuing attempt to curb grade inflation by a Duke University faculty who developed a grade-adjustment algorithm is now likely, with other schools about to adopt the system (see *The Chronicle of Higher Education* (1997, September 26).

[BACK to document]

43. One student writer went so far as to say, "We therefore suggest a boycott of the 1995 student/teacher evaluations. This boycott will provide a more effective means of communication than anything written on the evaluation itself. Something must be done about the trend of grade inflation. We as students refuse to contribute to the downfall of academia (Stern and Flynn, 1995).

[BACK to document]

44. It should go without saying that not all students are the same. SEF vary by maturity, and intellectual level, i.e., graduate student evaluations v. undergraduate (See, Divoky and Rothermel, 1988; Dilts, Samavati, Moghadam, and Haber, 1993), and therefore probably by campus and program. This may account in part for some of the wide variation in faculty attitudes toward SEF.

[BACK to document]

45. Some colleges count these remedial courses in the total number of credits required to graduate, thus further lowering the standards of the college degree. In terms of student ability and admissions standards, some data suggest that many college introductory text books have lowered their reading level from a grade 15 level (from 30 years ago) to about grade 11 reading level (presumably the reading ability of an 11th grade student. As measured by the Iowa Silent Reading Test, on some 4-year college campuses 23% of freshman are functionally illiterate independently reading at or below a 5th grade level.

[BACK to document]

46. Visit just about any World-Wide Web Listserv discussion group by academics. The membership of these Listserves are often a cross section of faculty experience with student abilities and expectations within a discipline. See what many of the contributors to these discussion groups relate

about their experiences with student expectations and behavior in the classroom.
[BACK to document]

47. A former law student from Duquesne University who was unable to achieve a B average filed suit charging that the school's grading policy constituted breach of contract. The suit contended that she had not been able to maintain the required 2.975 grade-point average, due to a new grading policy. The policy requires that not more than 30 % of students in a course can receive As and no more than 40% are to receive Bs. While newly enforced, the policy has existed for 20 years. A Duquesne official said. "It's a bell curve...There's no doubt about that" (*The Chronicle of Higher Education* (1997, September 12). One can seriously wonder if most psychometricians would call 70% As and Bs a bell curve.
[BACK to document]

48. Once again, I do not see this as basically a student problem; neither is it essentially a teacher problem, though these are all variables. The problem is a sociocultural one. To some degree, the problems is also a legal one. Faculty and administrators are perhaps overcautious in demands so as to avoid law suits. The question is, given the lowered admission standards and the above student characteristics, is it still possible to teach anything other than the "headlines" of a subject in a course? Is it possible to still teach the subtleties and the nuances of reasoning that are required to understand a subject matter?
[BACK to document]

49. See the recent *American Psychologist* that devoted a special section for SEF articles. Greenwald, 1997; Greenwald and Gillmore, 1997; d'Apollonia, and Abrami, 1997; Marsh, and Roche 1997; McKeachie, 1997.
[BACK to document]

50. See also Haskell, (1997c) section on: Assumption # 2: Statistical Significance of SEF of Teaching Effectiveness Measures Appropriate Learning.
[BACK to document]

51. See particularly section 1.3.3. The Role of Case Law for Kaplin and Lee's historical view of the relation between legal rulings on the elementary, secondary, and higher education levels.
[BACK to document]

52. It seems to be virtually unnoted that higher education has been uncritically transferring research findings and teaching methods originally designed for elementary and secondary levels to the college classroom. With exceptions like McKeachie and associates (1986) at the University of Michigan, there has not been a great deal of research with college level students (learning disabilities, so-called learning styles, etc, notwithstanding). The question is, how appropriately are finding from elementary and secondary education transferred to higher education? The answer, of course, depends on ones view of education (see Haskell, 1997c, section Assumption # 2: Statistical Significance of SEF of Teaching Effectiveness Measures Appropriate Learning). Many faculty believe that just as we have devalued the high school diploma, so are we now doing to the four-year college degree.
[BACK to document]

53. While it is not encouraging to contemplate, for years it has been known that the general academic quality of those enrolling in teacher education programs in the U.S. is lower than that of students enrolled in other university programs (Reyes, 1987, p. 18). In addition, it has been suggested for some time that the level of cognitive or mental development of the U.S population, from tests based on the work of the developmental psychologist, Jean Piaget, that about 50% of the U.S. population, including freshman and sophomore college students---and especially those entering teacher education programs---fail to function at a formal operational stage of reasoning, remaining at the concrete operational stage (e.g., Long, McGrary and Ackerman, 1979; Shyers and Cox, 1978; see also Stone, 1996).
[BACK to document]

54. The pressure faculty feel is considerable. I personally recall a situation a number of years ago where a faculty (an ex priest) applying for full professor was caught changing the scores on his student evaluations.

[BACK to document]

55. *University of Wisconsin*: In another ruling, the Chancellor at the University of Wisconsin refused to release SEF, citing a statute that disallows personnel evaluations from being released to public view. Students took the chancellor to court. However, after being advised by the state's Attorney General, citing Wisconsin's open-records law, the University of Wisconsin's campus released SEF to the public. Both the student and faculty senates passed resolutions in support of the Chancellor's refusal, and the university's lawyer concurred. Despite these resolutions, the Attorney General disagreed, writing that "the requested records are public records and the University's stated reasons for withholding access do not outweigh the public interest in the records" (*Chronicle of Higher Education*, 1994a, 1994b).
University of Idaho: A legal ruling, cited on the World Wide Web site of the *Topical Interest Group: Assessment in Higher Education* (Evaluating teacher evaluations, 1996), notes that the University of Idaho also recently went to court over the issue of whether SEF can be published. The student newspaper initiated a lawsuit when it was refused access to SEF for publication. The legal question was whether SEF is protected under privacy rights by the Idaho Code. In a ruling that seems to strain logical credulity, the court ruled that since the University did not consider students as the general public, the University was not breaking the law by allowing students access to the evaluations. Further, the opinion of the court was that according to state law, teacher evaluations are not protected as part of personnel records.

[BACK to document]

56. Many faculty are not aware of the extent of the confidentiality of student information. For example, (i) student scores or grades cannot be posted publicly by name, social security numbers, or any other identifier that can be known by anyone except the instructor and student; (2) student papers or lab reports that have names and grades on them cannot be left in places that are accessible to others; (3) students may not have access to other students grades in a class, (4) faculty are not to request student information without a legitimate educational reason; (5) student grades or other educational information may not be shared with other faculty members unless the faculty has a specific legitimate reason to know; (6) libraries are apparently prohibited from revealing to instructors what students have read the course reading material that the instructor has specifically put on reserve in the library for students to read; (7) student grades or other educational information can not even be revealed to the parent of the student (who may be paying for the student's education) without written permission of the student. There are many other restrictions as well.

[BACK to document]

57. According to *Black's Law Dictionary* (1990), libel includes.

> A method of defamation expressed by print, writing, pictures, or signs. In its most general sense, any publication that is injurious to the reputation of another. A false and unprivileged publication in writing of defamatory material....which tends to blacken a person's reputation or to expose him to public hatred, contempt, or ridicule.... to degrade him in the estimation of the community, to induce evil opinion of him in the minds of right thinking persons, to make him an object of reproach, to diminish his respectability or abridge his comforts, to change his position in society for the worse, to dishonor or discredit him in the estimation of the public, or his friends and acquaintances, or to deprive him of friendly intercourse in society, or cause him to be shunned or avoided....Almost any language which upon its face has a natural tendency to injure a man's reputation, either generally or with respect to his occupation.

[BACK to document]

58. I would like to make it very clear that I am not against mass higher education. I am a product of

it. I had to begin my educational career at community college as, unlike today, no 4-year college would admit me. So, I am not an elitist who is against expanding the boundaries of higher education. The basic difference between then and now, however, is that after being admitted traditional standards were upheld.
[BACK to document]

59. See for example, http://cri.ensmp.fr/mesr/mesr.html#numbers, June 9, 1997.
[BACK to document]

60. See my section on "Faculty Complicity in Adaptation to SEF" in Haskell (1977a). In one case, Kaplin and Lee (1995, Sec. 3. 7.) note "The Court also found that the state's entire system of "intricate administrative machinery [was] a highly efficient in terrorem mechanism.... It would be a bold teacher who would not stay as far as possible from utterances or acts which might jeopardize his living by enmeshing him in this intricate machinery....The result may be to stifle 'that free play of the spirit which all teachers ought especially to cultivate and practice.'"
[BACK to document]

61. Recently, a group of college and university presidents proposed that faculty input into decisions be reduced. The commission's 50-page report, *Renewing the Academic Presidency: Stronger Leadership for Tougher Times* makes recommendations to presidents, professors, trustees, and public officials. Copies may be obtained from the Association of Governing Boards of Universities and Colleges, One Dupont Circle, Suite 400, Washington 20036 (Leatherman, 1996b).
[BACK to document]

62. Some studies suggest that younger faculty do not seem to be as concerned with the complex issues revolving around SEF and tenure. See Kolevzon (1981), Avi-Itzhak and Lya (1986), Leatherman, 1996). I would suggest that many younger faculty have not been appropriately socialized into the academic profession with its rich heritage, culture and norms. As off-campus programs and distance learning increase, so will being socialized into the profession likely to further decrease.
[BACK to document]

3. A review of the grade inflation literature clearly shows that at perhaps the majority of institutions, Cs exist only minimally, with Ds and Fs nearly nonexistent.
[BACK to document]

References

AAUP (1975). Committee C

Abrami, P.C. (1989). How should we use student ratings to evaluate teaching? *Research in Higher Education*, *30*, (2). 221-27.

Academe: Bulletin of the American Association of University Professors. (1997, May/June). *83* (3).

Adelman, C. (1996, October 4). The truth about remedial work. *Chronicle of Higher Education*. A56.

American Federation of Teachers. (1996, February 15). A system of high standards: what we mean and why we need it. Executive Council Report. [Available online] http://www.aft.org/NS/GetLHP?url=%2fhigstan.htm&terms=system,high,standard#NSFIRSTTERM

American Psychological Association Monitor. (1994, November). Can universities be liable for incompetent grads? 7.

Avi-Itzhak, T., & Lya K. (1986). An investigation into the relationship between university faculty attitudes toward student rating and organizational and background factors. *Educational Research Quarterly, 10*, 31-38.

Bauer, H.H. (1996). The new generations: Students who don't study (prepared for the symposium, "The Technological Society at Risk" Annual Meeting, AOAC International, Orlando (FL), 10 September 1996). Email hhbauer@vt.edu for a copy of the article.

Black's Law Dictionary (With Pronunciations Sixth Edition (1990). St. Paul, Minn.: West Publishing.

Blum, D. (1992, November 4). U. Professors urge higher admissions standards. *The Chronicle of Higher Education*. A32.

Brimelow, P. (1996, April 22). Devalued diplomas. *Forbes*.

Carley V. Arizona Board of Regents, 737 P.2d 1099 (Ariz. App. 1987).

Chatterley, L.J. & Peck, D.M. (1995). We're crippling our kids with kindness!! *Journal of Mathematical Behavior, 14*, 429-436.

Copeland, J.D., & Murry, J. W., Jr. (1996). Getting tossed from the ivory tower. *Missouri Law Review, 61*, 233-327.

Crumbley, L. (1996). Due process/defamation issues [Not to be construed as legal advice] *Society for A Return to Academic Standards*. [Available online] http.tamu.edu:8000/~crumble/sfrtas.html

Crumbley, L.D., & Fliedner, E. (1995). Accounting administrators' perceptions of student evaluation of teaching (set) information. *Manuscript, Department of Accounting*, Texas A&M University.

Damron, J.C.(1995). The three faces of teaching evaluation. See also Damron, J.C. (1996). Instructor personality and the politics of the classroom. *Manuscript, Douglas College*, New Westminster, British Columbia, Canada V3L 5B2. (Earlier Version Faculty Matters, June 1994, (No. 5, Pages 9-12) and *Update*, September, 1994 (The Newsletter of the Okanagan University College Faculty Association). [Available online] at http://vax1.mankato.msus.edu/~pkbrandon/Damron_politics.html

Dewey, J. (1976). Academic freedom. In *John Dewey: The Middle Works*, 1899-1924. (pp. 53-56) Carbondale: Southern Illinois University Press. [Originally published in *Educational Review, 23*, 1902, 1-14]

DiBona v. Matthews, 269 Cal. Rptr. 882 (Cal. Ct. App. 1990).

Dilts, D.A., Samavati, H., Moghadam, M.R., & Haber, L.J. (1994). Student evaluation of instruction: Objective evidence and decision making. *Journal of Individual Employment Rights, 2*, 73-86.

Divoky, J.J., & Rothermel, A. (1988). Student perception of the relative importance of dimensions of teaching performance across type of class. *Educational Research Quarterly, 1*, 40-45.

d'Apollonia, S. & Abrami, P.C. (1997). Navigating student ratings of instruction. *American Psychologist, 52*. 1198-1208.

EEOC V. Franklin & Marshall College, No. 84-1739, United states court of appeals for the third circuit, 775 F.2d 110; 3 Fed. R. Serv. 3d (Callaghan) 282; 85 A.l.r. Fed. 669; 39 Fair Empl. Prac. Cas. (Bna) 211; 38 Empl. Prac. Dec. (Cch) P35,644, August 5, 1985, Argued, October 21, 1985, Decided.

Evaluating Teacher Evaluations (1996). Article in the topical interest group: Assessment in higher education submitted by glen9579@novell@uidaho.edu on 11/12/96. [Available online] http://marsquadra.tamu.edu/TIG/GeneralArticles/evaluatingteacherevaluations.ht

Fighting grade inflation. (1994, May 27) *Science*, *264*, 1255.

Furedy, J. (1995). Academic freedom, opinions and acts: The voltaire-mill perspective applied to current Canadian cases. *Univ. of New Brunswick Law Journal*, *44*, 131-4. Furedy@psych.utoronto.ca

Gold, J.J. (1997, September 12). The visiting professor: Student evaluations deconstructed. *The Chronicle of Higher Education*, B8.

Goldman, L. (1993, Spring). On the erosion of education and the eroding foundations of teacher education. *Teacher Education Quarterly*, *20*, 57-64.

Gordon, C., Hartigan, P, & Muttalib, N. (1996, November 22). Developmental learning at four-year colleges. *The Chronicle of Higher Education*. [Available online] http://chronicle.com

Gose, B. (1997, July 25). Efforts to curb grade inflation get an F from many critics. *The Chronicle of Higher Education*. A41. .

Greenwald, A.G. (1997). Validity concerns and usefulness of student ratings of instruction. *American Psychologist*, *52*. 1182-1186.

Greenwald A.G. & Gillmore, G.M. (1997). Grading leniency is a removable contaminant of student ratings. *American Psychologist*, *52*. 1209-1217.

Griggs v. Duke Power Co., 401 U.S. 424 (1971).

Guernsey, L. (1996, October 24). Remedial courses are widespread at American colleges, report says. *The Chronicle of Higher Education*. A8.

Hamilton, N.W. (1997, May/June). Peer review: The linchpin of academic freedom and tenure. *Academe: Bulletin of the American Association of University Professors*, *83*(3), 15-19.

Haskell, R.E. (1997a). Academic freedom, tenure, and student evaluations of faculty: Galloping polls in the 21st century. *Educational Policy Analysis Archives*, *5*(6). [Available online] http://olam.ed.asu.edu/epaa/v5n6.html

Haskell, R.E. (1997b). Abridgement of academic freedom, promotion, reappointment and tenure rights by the administrative use of student evaluation of faculty: (Part II) Views from the court. *Education Policy Analysis Archives*, *5*(6). [Available online] http://olam.ed.asu.edu/epaa/v5n17.html

Haskell, R.E. (1997c). Academic freedom, promotion, reappointment, tenure and the administrative use of student evaluation of faculty (SEF): (Part III) Analysis and implications of views from the court in relation to accuracy and psychometric validity. *Education Policy Analysis Archives*, *5*(6). [Available online] http://olam.ed.asu.edu/epaa/v5n18.html

Haskell, R.E. (1997d). Commentary: On Michael Theall's (and implied et al.) "A reply to Haskell and to Stake." *Educational Policy Analysis Archives*, *5*(8). [Available online] http://olam.ed.asu.edu/epaa/v5n8c3.html

Heller, S. (1986, December 17). Proper use of student evaluations at issue in professor's legal appeal to regain job. *The Chronicle of Higher Education*, p. 14.

Husbands, C., & Fosh, P. (1993). Student's evaluation of teaching in higher education: Experiences from four European countries and some implication of the practice. *Assessment & Evaluation on Higher Education*, *18*, 95-114.

Hertling, J. (1996, November 22). Education's growth in Hong Kong raises concern over quality. *The Chronicle of Higher Education*. A39.

Kaplin, W.A., & Lee, B. (1995). *The law of higher education: A comprehensive guide to legal implications of administrative decision making* (3rd Ed). San Francisco: Jossey Bass.

Keyishian v. Board of Regents, 385 U.S. 589, 604 (1967).

Knapper. C.K. (1977). Teaching evaluation and academic freedom. In C. Knapper, G. L. Geis, C. E. Pascal, and B. M. Shore (Eds.), *If Teaching Is Important ....: The Evaluation of Instruction in Higher Education*. (pp. 193-203). Toronto: Clarke, Irwin & Company, in association with the Canadian Association of University Teachers,

Kolevzon, M.S. (1981). Grade inflation in higher education: A comparative study. *Research in Higher Education*, *15*, 195-212.

Lambert, C. (1993, May/June). Desperately seeking summa. *Harvard Magazine*, 36-40.

Lansdale v. Tyler Junior College, 470 F.2d 659 (5th Cir. 1972).

Leo, J. (1996, September 16.). No books, please; We're students. *US News and World Report*.

Lasch, C. (1979). Schooling and the new illiteracy (Chapter Six). In *The culture of narcissism: American life in an age of diminishing expectations*. New York: Warner Books.

Leatherman, C. (1996, October 25). More faculty members question the value of tenure. *The Chronicle of Higher Education*, p. A12.

Leatherman, C. (1996b, September 13). Commission recommends strengthening college presidents' power. *The Chronicle of Higher Education*. A43.

Long, H.B., McGrary, K., & Ackerman, S. (1979). Adult cognition: Piagetian based research findings. *Adult Education*, *30*, 3-18.

Lovejoy, A. O. (1937). Academic freedom. *Encyclopedia of the Social Sciences*, (ed.), E.R.A. Seligman. New York: Macmillan.

Lovelace v. Southeastern Massachusetts University, 793 F.2d 419 (1st Cir.1986).

Magner, D.K. (1995, May 19). Mid-semester removal of professor roils University of Montana. *The Chronicle of Higher Education*, p. A25.

Mansfield, H.C. (1993, March 31). Harvard boils over article on grade inflation and race. *The Chronicle of Higher Education*. A27.

Marsh, H.W., & Roche L.A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias and utility. *American Psychologist, 52*. 1187-1197.

May, H. (1996, August 16). Ex-students sue universities over quality of education. *The Chronicle of Higher Education*. A29. [Available online] http://chronicle.com/che-data/articles.dir/art-42.dir/issue-49.dir/49a02901.htm

McKeachie W.J. (1997). Student ratings: The validity of use. *American Psychologist, 52*. 1218-1225.

McKeachie, W.J., Pintrich, P.R., Lin, Y.G., & Smith, D.A.F. (1986). *Teaching and learning post in the college classroom, A review of the research literatrure*. University of Michigan, National Center For Research To Improve Secondary Teaching and Learning.

McMurtry, J. (1991). Education and the market model. *Journal of Philosophy of Education, 25*, 209-217.

Menand, L. (1996). (Ed.), *The Future of Academic Freedom*. Chicago: University of Chicago Press.

Menand, L. (1993, May/June). The future of academic freedom. *Academe*. 11-17.

Millikan v. Board of Directors of Everett School District No. 2, 1980;

Morrow. G, & Sills, D.L. (eds.), (1968). Academic freedom: *International encyclopedia of the social sciences*. New York: Macmillan and The Free Press. 4-10.

Neath, I. (1996). How to improve your teaching evaluations without improving your teaching. *Psychological Reports, 78*, 1363-1372.

Parate v. Isibor, 868 F.2d 821 (6th Cir. 1989).

Pennsylvania State University (1996). *Policy On Confidentiality Of Student Records*. [Available online] gopher://info.psu.edu

Kaplin, W. (1997, April). Personal communication.

Power, B. M. (1997, June 20). The danger inherent in abusing academic freedom. *The Chronicle of Higher Education*. A52

Rebell, M.A. (1990). Legal issues concerning teacher evaluation. In Millman, Jason& Darling-Hammond, Linda, (Eds.), (pp. 337-355), *The New Handbook of Teacher Evaluation*. Beverly Hills, CA: Sage Pub.

Reisman, D. (1981). *On higher education: The academic enterprise in an era of rising student* *consumerism*. San Francisco: Jossey Bass

Regents of the University of California v. Bakke, 438 U.S 265, 322, 98 S.Ct 2733, 2759-2760, 57 L.Ed.2d 750 (1978).

Reyes, D.J. (1987, Mar/Apr). Cognitive development of teacher candidates: An analysis. *Journal of Teacher Education*. 18-21.

Riggin v. Board of Trustees of Ball State University, (1986).

Robert Kramer v. The President of the University of British Columbia, (1992).

Sacks, P. (1996). *Generation X Goes To College. An Eye Opening Account of Teaching in Postmodern America*. LaSalle, Il: Open Court. Power, 1997.

Scriven, M. (1996). The treason of the intellectuals. Adult Assessment Forum, 6 (2). [Available online] http://www.intered.com/treason.htm

Shattuck, R. (1997,July 18). From school to college: We must end the conspiracy to lower standards. *The Chronicle of Higher Education*. B6.

Shea, C. (1994, January 5). The subtleties of grade inflation students are said to desert the sciences in favor of easy grades in humanities. *The Chronicle of Higher Education*. A45.

Shyers, J., & Cox, D. (1978). Training for the acquisition and transfer of the concept of proportionality in remedial college students. *Journal of Research in Science Teaching*, *15*(1), 25-36.

Simon, W.E. (1996, March 19). The dumbing down of higher education. *Wall Street Journal*.

State Board for Community Colleges and Occupational Education v. Olson, (1984).

Stern, J., & Flynn, P.D. (1995, Feb 20). Students propose a course of action for grade inflation. *The Bucknellian*. [Available online] http://www.bucknell.edu/bucknellian/sp95/03-02-95/ops/4165.html

Stake, J.E. (1997). Response to Haskell: Academic freedom, tenure, and student evaluations of faculty. *Educational Policy Analysis Archives*, 5. [Available online] http://olam.ed.asu.edu/epaa/v5n8.html

Stedman, L.C. (1996). The achievement crisis is real: A review of the manufactured crisis. *Education Policy Analysis Archives*. [Available online] http://olam.ed.asu.edu/epaa/v4n1.html

Stichler, R.N. (1997, May/June). Academic freedom and faculty responsibility in disciplinary procedures. *Academe: Bulletin of the American Association of University Professors*, *83*(3), 20-22.

Stone, J.E. (1996). Developmentalism: An obscure but pervasive restriction on educational improvement. *Education Policy Analysis Archieves*, *4*(8) [Available online] http://olam.ed.asu.edu/epaa/v4n8.html

Stone, J.E. (1995). Inflated grades, inflated enrollment, and inflated budgets: An analysis and call for review at the state level. *Education Policy Analysis Archives*, *3*(11) [Available online] http://olam.ed.asu.edu/epaa/v3n11.html

Strosnider K. (1996, December 6). Court orders students who dropped a course to pay for it. *The Chronicle of Higher Education*. A56. [Available online] http://chronicle.com

Summerville, R.M., et al. (1990). Grade inflation: The case of urban colleges and universities. *College Teaching*, *38*(1), 33-38.

Sunstein, C.R. (1996). Academic freedom and law: Liberalism, speech codes and related problems. In Menand, L. (Ed.), *The Future of Academic Freedom*. (pp.93-118) Chicago: University of Chicago Press.

Sykes, C.J. (1988). *Profscam: Professors And The Demise Of Higher Education*. Washington, D.C.: Regnery Gateway

Theall, M. (1997). On drawing reasonable conclusions about student ratings of instruction: A reply to Haskell and to Stake. *Educational Policy Analysis Archives*, *5*(8). [Available online] http://olam.ed.asu.edu/epaa/v5n8c2.html

The Chronicle of Higher Education. (1991, January 16). SAT's better freshman predictor than grades. A35.

The Chronicle of Higher Education. (1997, August 8). Critics of SAT and ACT hail decline in colleges that use them. A41.

The Chronicle of Higher Education. (1997, September 12). Former Duquesne law student sues over grade curve. A13.

The Chronicle of Higher Education. (1997, September 26). The controversial plan to curb grade inflation at Duke University is back. Section: Notebook. A46.

The Chronicle of Higher Education. (1997, November 14). A professor's personal teaching style wins him praise and costs him his job. A12

Trotter, B. (1977). The teacher and the goals of the university. In C. Knapper, G. L. Geis, C. E. Pascal, & B. M. Shore (Eds.), *If Teaching Is Important ....: The Evaluation of Instruction in Higher Education*. (pp. 150-158). Toronto: Clarke, Irwin & Company, in association with the Canadian Association of University Teachers.

Trout, P.A. (1997, Oct). What the numbers mean providing a context for numerical student evaluations of courses. *Change: The Magazine of Higher Learning*, *29*, 5, 25-30.

University of Michigan. (1994, Dec 12). Toward a definition of tenure. *Senate Advisory Committee on University Affairs Standing Subcommittee on Tenure*. Endorsed Unanimously by the University of Michigan Senate Assembly. [Available online] http://www.umich.edu/~aaupum/tenure05.htm

University of Idaho: A legal ruling, cited on the World Wide Web site of the Topical Interest Group: Evaluating Teacher Evaluations (1996). Article in the Topical Interest Group: Assessment in Higher Education submitted by glen9579@novell@uidaho.edu on 11/12/96. [Available online] http://marsquadra.tamu.edu/TIG/GeneralArticles/evaluatingteacherevaluations.html

Walker, D. (1992, October 14). British universities urged to shun remedial courses. *Chronicle of*

Higher Education. [Available online] http://chronicle.com

Weeks, K. (1988). Student evaluations of faculty and the legal context. *Lex Collegii*. 11 (Winter).

William Sypher v. Vermont State Colleges Faculty Federation. (1982). 5 VLRB 102.

Williams, W.M., & Ceci, S.J. (1997, October) How'm I doing? Problems with student ratings of instructors and courses. *Change: The Magazine of Higher Learning, 29*, 5, 13-23.

Winston, G.C. (1997, October). Why can't a college be more like a firm? *Change: The Magazine of Higher Learning, 29*, 5, 33-38.

Zirkel, P.A. (1996). *The law of teacher evaluation*. Bloomington, IN: Phi Delta Kappa Educational Foundation.

About the Author

Robert E. Haskell

Brief Bio ~ Robert E. Haskell has been teaching college and university level courses for over twenty years. He earned his Ph.D. from the Pennsylvania State University in Psychology and Social Relations, his M.A., and B.A. from San Francisco State University. His areas of research and teaching include: transfer of learning, analogical reasoning, small group dynamics. Major publications include: four books, the latest of which is, *The Future of Education and Transfer of Learning: A Cognitive Theory of Learning and Instruction For The 21st Century* (forthcoming), and numerous presentations, chapters, and research articles in national and international journals. He also serves on several editorial review boards, and is Associate Editor of The Journal of Mind and Behavior. He is former Chair, and currently Professor of Psychology, Department of Social and Behavioral Sciences, University of New England.

Professor of Psychology
University of New England
Biddeford, Maine 04005

UNE Home Page: http://www.une.edu/sbs/haskell.html

E-mail: rhaskell@javanet.com

Copyright 1997 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://olam.ed.asu.edu/epaa

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, Glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692)

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Marshall University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
Arizona State University

Dennis Sayers
University of California at Davis

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina--Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher Education

Mary P. McKeown
Arizona Board of Regents

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois--UC

Robert T. Stout
Arizona State University

---

archives | abstracts | editors | board | submit | comment | subscribe | search

---

This article has been retrieved 891 times since December 9, 1997.

Education Policy Analysis Archives

Volume 5 Number 22 December 9, 1997 ISSN 1068-2341

A peer-reviewed scholarly electronic journal.
Editor: Gene V Glass Glass@ASU.EDU.
College of Education
Arizona State University,Tempe AZ 85287-2411
Copyright 1997, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

Changing Definitions and Off-loading Responsibility in Alberta's Post-Secondary System

Robert J. Barnetson
Alberta College-Institutes Faculties Association
Alberta, Canada

Abstract

The introduction of a performance-based funding mechanism by Alberta's provincial government alters the public definition of "educational quality" and fully shifts the responsibility for declining educational quality from the provincial government onto institutions. This article outlines the process by which the provincial government has compelled institutions to accept this redefinition and transfer despite the substantial loss of institutional autonomy it entails. The implications of this change are explored and possible reasons are suggested.

Introduction

The introduction of performance indicators (PIs) in Alberta's performance-based funding mechanism has changed the definition of education quality and fully shifted the responsibility for providing a high-quality post-secondary education onto institutions. This ignores the pivotal role government funding plays in maintaining educational quality. One result of this process is a significant loss of institutional autonomy. A second result has been the imposition of a market model on education that shifts responsibility for education from government to individuals.

Background

Alberta's public service has recently undergone a massive downsizing as the provincial government has eliminated its deficit through reductions in public spending. The decline in government funding to the post-secondary education (PSE) system totaled 21% between 1994 and 1997 (AECD, 1994). When coupled with spiraling enrollments and stagnant provincial funding, Alberta's 21 public colleges, universities and technical institutes have seen per-student, constant-dollar funding fall from $6165 in 1980/81 to $3267 in 1996/97--a reduction of 47% in 17 years. Funding and enrollment projections through 2005 suggests this pattern will hold (AAE, 1992; Treasury, 1997).

This trend in post-secondary funding has occurred worldwide and concurrently with the contraction of the welfare state. This suggests a return to the classically liberal approach to government where liberty is defined as "freedom from restraint" in contrast to the reform liberal definition of liberty as the "freedom to act" (Gibbins and Youngman, 1996). This shift has been caused (or, alternately, legitimized) by large government deficits and debts.

Debt and deficit; revenue and expenses

Upon election in 1993, Alberta's government faced a substantial deficit and growing debt--the legacy a decade of upheaval in oil and gas prices--and set about attempting to reduce government expenditures. Government rhetoric about spiraling public-sector costs as the prime cause of both the debt and deficit, according to Kevin Taft, is largely inaccurate: overall government spending had been declining since 1986.

> "...(former premier Don) Getty's government had kept this decline fairly quiet, still wanting Albertans to believe they were getting 'the best' from their province, an expectation strongly engendered during the Lougheed years. As a result, the public was still under the misconception that Alberta spent far more than other provinces."

The public belief that programs were still rich, reinforced by a strong public opposition to tax increases left open a whole new strategy for (current) Premier (Ralph) Klein's incoming government. He and his ministers strongly reinforced the mistaken perception that spending was out of control and argued vigorously to cut expenditures (Taft, 1997, p. 12). Between 1986 and 1992, Getty's cuts saw government spending fall by 15% (adjusted for inflation) and, by fiscal 1991/92, Alberta spent $4593 per capita on public services as compared to the Canadian average of $4758 (McMillian and Warrack, 1995). Yet, in June 1994, Klein told the Edmonton Journal, "When our government took over a year and a half ago, we saw uncontrolled spending" (Taft, 1997, p. 25).

Cooper and Neu (1995) suggest that the real reason that expenditures had outstripped revenues during the Getty years was that the government was not only feeling the effects of the low resource prices, but also the effects of granting the oil-and-gas lobby tax breaks as oil prices fell. This means that the government received lower royalties because of lower prices and also absorbed a portion of the oil-and-gas companies' losses (due to these same lower prices) by accepting a lesser percentage of the sale price in royalties.

McMillian and Warrack (1995) note that the government had three way to compensate for its deficit: (a) reducing public expenditures, (b) increasing revenues or (c) a combination of reducing expenses and increasing revenue. The decision to concentrate on the expense side of the government balance sheet (to preserve the Alberta Advantage of low taxes) was a conscious decision (as opposed to being the necessity Klein often portrays it as) and reflects the political difficulty of selling tax increases to Albertans used to low taxes.

The Alberta Advantage: Rhetoric or Reality?

The Alberta Advantage is premised on "a powerful combination of low taxes, fiscally responsible government, abundant natural resources and a well-educated workforce" (Klein, 1996). The Klein government's belief that low taxes stimulate business and consumer spending explains why it has focused solely on reducing expenditures to balance the budget rather than also attempting to increase revenues by raising taxes.

The dogmatic belief that low taxes attract investment in Alberta by companies runs contrary to conventional approaches to estimating the profitability of locating businesses in a region (Drugge, 1995). The corporate sector estimates profitability based upon the possibility of high-volume sales combined with access to low-cost, high-quality inputs and inexpensive transportation. Further, evidence gathered over the past 15 years clearly demonstrates that a low-tax approach to economic development (the basis of the Alberta Advantage) has not been successful in Alberta.

McMillian and Warrack (1995) note that Alberta's taxes were lower than any other province prior to Klein's election (approximately 75% of the Canadian average) and had been kept this low by the presence of natural resources revenues for the past decades. (In 1993/94, Alberta received approximately 40.3 per cent of its revenue from individual and corporate taxes while the Canadian average was 61.3 per cent. The difference was largely made up of the taxes assessed on the extraction of natural resources.) If low taxes do stimulate economic growth in Alberta, there should be evidence in Alberta's growth through the 1980s.

Between 1981 and 1993, Alberta's average annual growth in retail sales was 0.3 per cent, substantially lower than that of higher-tax British Columbia (1.3 per cent) and lower than the Canadian average of 0.7 per cent (Drugge, 1995). Average annual real business investment in Alberta during this period (investment in plants, machinery and equipment) was -3.0 per cent, while in British Columbia it was 1.4 per cent and the Canadian average was 2.9 percent. The average annual real growth rate in Alberta from 1981 to 1993 was 0.6 per cent. British Columbia had a 2.9 per cent growth rate and the Canadian average was 2.5 per cent.

What this analysis demonstrates is that Alberta's low-tax strategy has been unsuccessful in stimulating growth. This is because Alberta's economy is largely tied to the natural resources market and a low-tax strategy lacks the power to successfully counter the fluctuations in the resource market. The Alberta Advantage does not work in Alberta.

In 1990/91, Alberta personal and corporate taxes equaled $2885 per capita (McMillian and Warrack, 1995). The Canadian average was $3681. Additionally, Alberta has a tax capacity approximately 33 per cent greater than the Canadian average, meaning having taxes comparable to the rest of Canada would have actually netted the government $4331. Given that low taxes have little ability to stimulate growth in a resource- dominated economy, raising them towards the Canadian average in conjunction with reductions in public spending provides a balanced solution to the revenue problem without unduly burdening taxpayers. This solution has been successfully implemented in the neighboring province of Saskatchewan.

Further, by adjusting both revenues and expenditures, Alberta could have maintained its investment in human capital that business owners rank as vastly more important in business success than low taxes. The assault on Alberta's investment in

> "human capital negatively affects precisely those features such as quality of labour and management noted as the key resources of economic success in the small business surveys cited above. These expenditure reductions must therefore raise serious questions about

> the underlying logic of the government's policies." (Drugge, 1995, p. 189).

The importance of resource revenue in Alberta's budget further suggests that spending on public programs wasn't the source of the deficit but that low tax revenues were. And attempting to control the deficit solely by cutting public-sector spending in order to avoid tax increases makes no sense. That the government did not take note of the economic lessons of the Getty years (because of its fixation on the politically saleable Alberta Advantage) and drastically down-sized the public sector at the expense of the province's economic future, suggests either incompetence or priorities other than ensuring the economic success of every Albertan through access to public services like post- secondary education.

Per-student funding and the changing role of government

Returning to the earlier discussion of the effect of ideological change on the degree of responsibility assumed by society for providing a post-secondary education system, the disaggregation of responsibility for the reform liberalist "freedom to act" can be seen in the government's human resource strategy, People and Prosperity:

> Continuous learning and the updating of skills is a shared responsibility. The primary onus is on individual Albertans, but strategies are needed to help them access learning opportunities and obtain the skills and knowledge they need to be successful. Student assistance ensures that financial barriers do not act as a deterrent to Albertans pursuing adult learning. Alberta's schools, universities, colleges and technical institutes play a key role in our human-resource strategy. Schools are responsible for providing education programs that develop individual potential and prepare young Albertans for daily living, the world of work and lifelong learning. Adult learning institutions have a responsibility to provide high- quality, accessible learning opportunities to people who are preparing for careers and to those who wish to update their skills. Employers, employee groups and unions have a responsibility to facilitate learning opportunities in the workplace (Government of Alberta, 1997a, p. 10).

Notably absent is the government's share of responsibility for continuous learning and the updating of skills. Implicitly, this excerpt suggests that the government is responsible for providing student assistance to ensure accessibility, however, student assistance is increasingly in the form of loans (which are now administered by banks) thus accessibility is being bought by Albertans at the cost of assuming large personal debts. It could also be argued that the government is implicitly agreeing to fund schools and adult-learning institutions. But, as noted above, per-student government grants have been in a period of long-term and significant decline.

Educational quality and the implications of funding reductions While defining quality is difficult, Harvey and Green (1993) outline five broad

660

approaches to the concept:

1. The classical view of quality as transformation sees it in terms of change from one state to another (i.e. a high-quality PSE would transform a student into a different person).
2. The exceptional perspective sees quality a something special and is closely associated with the idea of excellence attainable by very few;
3. Quality as perfection equates quality with a consistent or flawless outcome;
4. The fitness for purpose approach defines quality as meeting the needs of the customer or end-user; and
5. Quality as value for money emphasizes quality as the return on investment.

The traditional view of quality (i.e. the transformative ability of education) is most consistent with Canadian higher education. The purpose higher education has been to develop students' knowledge, skills, and attitudes and the process (involving introspection, synthesis and integration) has varied according to the experience, aptitude and motivation of the students. This differs from market-based concepts of quality in that something is being done to the consumer as opposed to something being done for the consumer. By changing the consumer (and, presumably, the consumer's references) during the process, higher education violates the assumption of pre-existent purpose necessitated by the rationality inherent in market models of decision-making (Harvey and Green, 1993; Stone, 1988; March, 1976; March and Olsen, 1976).

With the classical definition of quality in mind, there are three preconditions for a high-quality post-secondary education system:

a) PSE must be accessible for students such that levels of tuition and net debt at graduation do not deter students from pursuing education and training according to their aptitude.
b). Public colleges, universities and technical institutions must be able to attract and retain high-quality teachers and researchers.
c) Institutions must be able to provide adequate infrastructure, including an appropriate physical plant and modern equipment and information resources for students.

Declining per-student funding has impaired the ability of Alberta's PSE system to provide a high-quality education.

Accessibility

In constant 1995 dollars, tuition and fees at The University of Calgary (one of Alberta's three research-focused universities) will have increased from $1326 in 1989 to $3837 in 1999 (UCSU, 1996). This 189% increase is mirrored at the Universities of Alberta and Lethbridge. College and technical-institute students have seen similar changes. For example, annual tuition at Grant MacEwan Community College rose 174% between 1989 and 1997, jumping from just over $500 to nearly $1700 while Red Deer College students will have seen tuition jump by

227% between 1989 and 1998.

Net student debt at graduation is also increasing with average student debt at graduation (of those students with loans) rising from $6076 in 1987 to $11,604 in 1995 (AAE, 1990; UCSU, 1996). University-student debt in 1996 averaged $15,518 while college students averaged $9172 and technical-institute students averaged $8752 (AECD, 1997e).

Statistics Canada data on university tuition and participation rates suggests that costs associated with post-secondary education have risen faster than the resources available to students (Little, 1997). From 1989 to 1994, the percentage of an average family's income that an average tuition assessment comprised rose from 3.1% to 4.9%. Nationally, between 1984 and 1995, tuition increased (in real dollars) by 75% while Canada Student Loan allotments increased by only 55%. The declining rates of employment of 20- to 24-year-olds (73% 1989, 65% in 1994) and stagnant wage levels mean that students may be leery of assuming long-term debt that they will be unable to manage upon graduation. For example, in 1993/94, 23% of students owing money of loans defaulted (Calgary Herald, 1994).

Many students groups suggest that high levels of tuition and debt are barriers to accessibility (ACTISEC, 1997). Their contention may have some statistical support in a recent Statistics Canada report that notes a decline in overall enrollment levels after 1993. While the percentage of 19- to 24-year-olds enrolled has continued to increase, this increase has slowed significantly (Little, 1997). A survey of high -school graduates in Alberta noted that 38% of graduates not planning on pursuing a post-secondary education were making that choice for financial reasons while 64% of respondents agreed that post-secondary education was getting too expensive (Krahn and Lowe, 1997).

High-Quality Teachers and Researchers

Alberta's colleges, universities and technical institutes will need to replace approximately 40% of their faculty over the next 10 years due to an uneven age distribution (AAS:UA, 1995). Low salaries as compared to other provinces and industries combined with poor working conditions will make this a challenging task.

In a 1985/86 ranking of 18 Canadian universities, mean salaries at the Universities of Alberta and Calgary placed second and fourth, respectively. By 1995/96, the U of A had slipped to 16th place while the U of C occupied 17th (Robb, 1997). Corrected for inflation, salaries at the U of A have actually declined slightly in the same time-frame (AAS:UA, 1996). The wage gap between positions in Alberta and positions in British Columbia and Ontario means that, assuming current conditions hold, faculty in Alberta earn nearly $200,000 less over the course of their life-time (CAUT, 1996). While wages for faculty across the country have stagnated or declined, Alberta's decline has been relatively more rapid. This has been exacerbated by declining infrastructure (see below).

College and technical-institute faculty face a similar wage-gap with their counterparts in Ontario and British Columbia earning 20-25% more at each step of the salary grid (personal communication, Evelyn Wieland, 25 June 1997; personal communication, David Piasta, 4 June 1997).

Colleges and technical-institutes also face substantial competition from other industries for the professionals who are potential faculty. Instructors at the Northern Alberta Institute of Technology can expect to make $40,000 per year but can command salaries of up to $80,000 in high-demand industries (Avery, 1997). This wage disparity resulted in, for example, the loss of the entire faculty complement in the province's only Crane and Hoist program to industry during the summer of 1997 (personal communication, Terry Sway, 13 September 1997).

Faculty throughout Alberta are also facing increasing workloads because of rising teacher-to-student ratios. At the University of Alberta, the teacher-to-student ratio increased by 20% between 1991 and 1995 (AAS:UA, 1996). At the Southern Alberta Institute of Technology, the ratio jumped by 88% between 1992 and 1995 (SAIT, 1996).

In addition to generating greater workloads, rising faculty-to-student ratios decreases the quality of education provided. Research in Canadian universities suggests that classroom technique is the key determinant of educational outcomes (Gilbert, 1995). This ignores that classroom technique is largely determined by class size: for example, it is easier to encourage discussion and interaction in a class of 30 than in a class of 100. Besides forcing changes to less effective instructional techniques, increasing class sizes alters evaluation from written exams (that test students' ability to apply their knowledge and skills to novel problems) to multiple-choice tests (that emphasize the ability to associate concepts and commit information to short-term memory).

Infrastructure

In 1993, the Ministry of Advanced Education and Career Development estimated the deferred infrastructure investment on its $4 billion capital stock to be approximately $400 million (AECD, 1993). Deferred investment in infrastructure encompasses every aspect of institutional infrastructure from building maintenance and replacement to renewing library and equipment stocks. This issue (despite a recent $105 million injection into the system by the government over three years) remains significant and impacts student learning and faculty recruitment (AECD, 1997a).

Performance-Based Funding in Alberta's PSE System

As part of the provincial government's 1994 White Paper on PSE, the government committed the Ministry to develop a reporting framework that assessed the outcomes of PSE in an effort to improve the accountability of the system for the funding it received. This was tied to a commitment to develop a performance-based funding mechanism designed to "reward performance and productivity in publicly supported post- secondary education" (AECD, 1994, p. 15). By late 1996, the goals for the funding mechanism had become "to reward performance and promote excellence" (AECD, 1996, p. 3)

While a substantial number of measures of accountability existed within the system prior to 1994, they largely focused on inputs and process (AECD, 1995). Specific criticisms of then-existing measures of accountability included:

a) a lack of systematic use;
b) inadequate information provision to prospective students about educational outcomes;
c) not providing basic information, such as cost information, at the program level; and
d) inconsistent inter-institutional data definitions impeding comparisons.

Institutional consultations developed a series of 76 of Performance Indicators (PIs) to meet the requirements of the accountability framework. Nine of these indicators are considered Key Performance Indicators (KPIs) and are utilized in a performance-based funding mechanism (four indicators apply only to universities' research mandate) that allots each institution an additional 1-2.5% of its operating grant based on its KPI scores (AECD, 1996).

Alberta performance-based funding mechanism

Alberta's performance-based funding mechanism (PBFM) takes institutional performance on each of a series of performance indicators and compares it to benchmarks. Points are awarded to an institution for its performance on each indicator, the overall institutional score is tallied and additional funding is awarded based on an institution's ranking against all other institutions' scores.

Using the enrollment indicator as an example, urban institutions that saw enrollment fall by more than -2% received zero points. Enrollment changes between -2% and 0% received 20 points. Enrollment increases between 0% and 4% received 25 points and enrollment increases greater than 4% garnered a full 30 points. Each institutions' score on the enrollment indicator would be added to its scores on the other indicators to create an overall assessment of performance that would then be compared against the scores of other institutions.

The nine indicators used fell into four categories: responsiveness, accessibility, affordability and (for universities only) research excellence (AECD, 1997b). What follows is a brief explanation of each group of indicators.

Responsiveness

The responsiveness of the system to changing economic, social and cultural needs is measured by two indicators: the employment rates of graduates and graduates' satisfaction with their program.

Accessibility

The indicator measuring accessibility focuses on institutions' ability to maintain and improve the enrollment levels. This indicator is a near-perfect (.95) predictor of overall KPI performance thus generates immense pressure for institutions to further increase enrollment.

Affordability

The two indicators measuring affordability examine the ability of institutions to minimize administrative costs and generate revenue from sources other than tuition from credit programs and government grants. No measure of affordability to learners is made.

Research excellence

Four indicators measure the performance of Alberta's three research-focused universities. Institutions' national peer-group ranking per full-time faculty member in terms of council monetary awards, citations per research publication, and community and industrial funding are measured, as well as sponsored revenues as a percentage of government operating grants.

Quality in Alberta's PSE System

Alberta's performance-based funding mechanism, with its mandate to "encourage and reward excellence", both redefines the public's notion of high-quality education and shifts the full responsibility for providing it on to individual institutions (AECD, 1996, p. 1).

Redefining Quality

Performance indicators draw on an industrial metaphor of production and focus on input-out analysis (Emberley, 1996). This approach assumes that (1) inputs (e.g. iron, wood, etc.) are uniform and passive and (2) outputs (e.g. beams, tables, etc.) are standardized therefore making it is possible to find a best (i.e. maximally effective and efficient) way to transform inputs into outputs by comparing different approaches in a purely statistical manner.

If the inputs and outputs are variable then a common metric by which to measure is necessary for PIs to make valid comparisons between processes. For example, the appropriateness of sports cars and sedans for various purposes-- with inputs (i.e. design) and outputs (i.e. performance) that vary substantially-- can be compared based on common metrics like fuel efficiency (km/l) and speed-time measures (kmh).

Education provides neither uniform inputs and outputs nor common metrics. Students and instructors (i.e. inputs) are neither passive nor uniform. Graduates (i.e. outputs) are similarly unique in the degree to which students' knowledge, skills and attitudes can be developed and for what ends. The process (involving introspection, synthesis and integration) is similarly variable according to students' experiences, aptitudes and motivation and resistant to universal performance metrics (although it may be possible to specify measures appropriate to narrow subfields along the lines of professional designation exams). Attempting to make valid assessments of performance based solely on statistical calculation will yield a less than optimal judgment.

In an attempt to get around the conceptual and operational difficulties posed by outcomes-based performance indicators, Alberta's government has focused on the more easily quantifiable economic outcomes and ignored education's social and cultural outcomes. This vast omission is dismissed with "the social and cultural outcomes, while important, have not been clearly articulated" (AECD, 1996, p. 6). Given the importance of the social and cultural outcomes of education, it is unlikely that the government would fail to articulate them without reason. One explanation for this omission is that intractable problems associated with developing quantitative PIs for social and cultural outcomes. Another explanation is that, from the perspective of the market model, education is solely an economic good thus measurement should focus on this aspect. By using quantitative measures of quality, KPIs shift the definition of quality from that of transformative to quality as value for money and as fitness for purpose.

The two outcomes of PSE in Alberta that are measured are the employment rate of graduates and the satisfaction of graduates with their program. The satisfaction of graduates with their program suggests a fitness for purpose orientation to quality: is the customer satisfied? Measuring employment rates is yet another measure that reflects the fitness for purpose approach to quality: if one goal of (most) students is to be employed (Barnetson, 1997), the employment rates of graduates are simply a proxy measure for customer satisfaction.

It is the very idea of performance-based funding that suggests a value for money approach to quality. According to the government, "a performance-based funding approach means Albertans are getting maximum returns on their investment in post- secondary education" (AECD, 1997c, p. 7).

By emphasizing customer satisfaction and the economic success of graduates, the government has subtly redefined the concept of quality such that it excludes traditional indicators of quality. The declines in accessibility, quality of faculty, and infrastructure that are the direct result of the long-term reduction in per-student funding from the provincial government are completely ignored by KPIs and the government can thus neatly side-step criticism (based on these grounds) that the quality of post-secondary education has declined.

Further, by selecting measures such as satisfaction and the employment rates of graduates as indicators of quality--indicators that institutions historically have done well by--the government creates data that justifies its substantial reduction to post- secondary funding. The high score of almost every institution on every indicator seems to support the government's contention that the post-secondary system is as good (if not better) than it was before the 21% reduction to its funding between 1994 and 1997.

During the press conference announcing the performance awards, Minister for Advanced Education and Career Development Clint Dunford explained, "We believe that striving towards the goals will promote continuous improvement at individual institutions and throughout the system. Based on our results, thus far, excellent progress is being made" (AECD, 1997d, p. 2). That the lack of baseline data makes the contention of progress impossible to prove or disprove simply gets ignored in light of high levels of institutional performance. Also glossed over is that the level of institutional performance was guaranteed by government selection and pretesting: the government sets the levels of performance that are labeled as indicative of "excellence."

Off-loading Responsibility for Quality Education

By measuring each institution's performance on KPIs, the government subtly shifts the full responsibility for the provision of high-quality education onto the institution itself. Institutions are measured on their ability to turn out employable, satisfied graduates in increasing numbers. The nature of the measurement attributes the responsibility for an institution's performance to the institution when, in reality, the ability of an institution to improve its performance is largely constrained by environmental factors. Continually shrinking per-student government funding, declining granting council budgets and limited alternative revenue streams for institutions creates circumstances that impair institutions' ability to provide accessible programs staffed by high-quality teachers and researchers and supported by adequate infrastructure.

Again, the government neatly sidesteps criticism. Not only are institutions generally doing exceptional jobs according to the KPI performance evaluations, but those institutions that "require" improvement are assigned responsibility for it. According to the government brief on the KPI system, "There are areas for improvement and institutions will be expected to focus on those in the future" (AECD, 1997c, p. 7).

Loss of Institutional Autonomy

Government control of the post-secondary purse-strings is substantial and has both created a desperate need within institutions for additional funding (by reducing annual operating grants by approximately $250 million between 1994 and 1997) and proffered a partial remedy (though the $15-million performance-funding envelope). In order for institutions to receive the remedy, however, they must accept both the government's redefinition of quality (an unprecedented infringement on institutional autonomy that shifts considerable indirect control over program objectives and delivery methods from institutions to the government) and the transfer of responsibility for declining educational quality from the government (that created the initial decline in quality through long-term budget cuts) to the institution.

By initially acquiescing to the government's performance-based funding mechanism, institutions have surrendered a substantial amount of their autonomy in exchange for the return of less than 5% of the funding the government removed from the system between 1994 and 1997. Government is now free to change both the level of institutional performance it considers acceptable on any current indicators and introduce additional KPIs from the 67 other PIs it is collecting data on. This could include the introduction of an indicator stipulating mandatory completion rates (a KPI that was dropped from the performance-envelope calculation only weeks before the final tallying of institutional scores) that would pressure institutions to graduate students regardless of performance and infringe upon academic autonomy. Pressure has already been applied to faculty in Red Deer College's University transfer program who were told by their Board of Governors in the spring of 1997 to increase their completion rates (rates that were already equivalent to those of both of the province's major universities) or risk program closure (personal communication, Jim Scott, 1 June 1997)

This transfer of power has gone virtually unnoticed in both the media and institutions. For example, no concern has been raised regarding the overwhelming importance of the enrollment indicator on overall institutional KPI performance (as outlined above) despite the substantial pressure it exerts on institutions to increase enrollment without additional per-student funding increases to their base grant. The educational consequences of this (e.g. larger classes, fewer assignments and less instructor-student contact) are substantial and negative, however, ironically, this is considered an indicator of quality by the performance-funding mechanism.

Redefinition of purpose in education

The changes in how educational quality is defined and who is responsible for ensuring it created by KPIs suggests that there is a general redefinition of educational purpose occurring. The liberal-arts tradition that has driven post-secondary education is incompatible with government changes in health care, social services and other areas where citizens are being reframed as customers. The increase in government control over educational content and method of delivery facilitated by KPIs reduces the barriers the government faces in imposing a market model on the PSE system.

If students are thought of as consumers of education, the important concept of entitlement is removed from the debate (personal communication, Raj Pannu, 12 September 1997). A consequence of this change is the elimination of the government's responsibility for providing education: education is a commodity (to be purchased by the individual) rather than a right (to be provided by society).

Cast as an economic good, education then has value only in economic terms. The most apparent economic outcome of increasing levels of education is increasingly favorable employment outcomes thus education is valued solely in terms of its return on investment and becomes indistinguishable from one of its outcomes: job training. As an economic good, there is no value placed on social or cultural outcomes; indeed, there isn't even a framework in which to discuss them as anything more than trivia byproducts of a fundamentally economic process.

If education is framed as job training (which emphasizes the accrued personal benefits of higher salaries and levels of employment and downplays the societal benefits of higher quality of life), responsibility for obtaining education (and overcoming the barriers to obtaining it) shifts onto the individual. This disaggregation of responsibility (when carried out across all sectors of government) creates an environment whereby competition between individuals replaces cooperation as the basis of society (Kohn, 1992). This approach ignores that cooperation is the fundamental assumption of society and that competition disproportionately favors those with greater personal resources. This same group is further and differentially benefited by the tax reductions made possible by lower government expenditures on education and other public services (McQuaig, 1996).

Conclusions

The quality of education (i.e. its transformative ability) provided by Alberta's post- secondary system is declining as a result of long-term declines in per-student government funding. The government's KPI-based funding mechanism changes the definition of education quality which obfuscates public discussion and thereby deflects criticism. At the same time, this mechanism off-loads the responsibility for providing a high-quality post-secondary education to institutions and ignores the instrumental role of government funding.

Institutions have been compelled to accept this change by the financial power of the government--the same power that created the initial decline in educational quality--and as a result have surrendered substantial institutional autonomy. The loss of institutional autonomy facilitates the imposition of a market model into the education sector. The market model sees education in terms of its economic outcomes thus redefining education as job training. This shifts responsibility for education from government to the individual thereby engendering competition between differentially advantaged consumers. At the same time, it trivializes debate about the social and cultural outcomes of education because the market model cannot accommodate serious discussion of non-quantifiable outcomes.

References

AAE. (1990). Tuition fee policy review: A discussion paper. Edmonton, AB: Government of Alberta, Alberta Advanced Education.

AAE. (1992). High-school graduates and enrollment demand. Edmonton, AB: Government of Alberta, Alberta Advanced Education.

AAS:UA. (1995). "Faculty renewal and early retirement." Association of academic staff: University of Alberta newsletter. Issue 4. pp. 8-9.

AAS:UA. (1996). "Operation renewal: The salary challenge." Association of academic staff: University of Alberta newsletter. December. Issue 9. pp. 1-6.

ACTISEC. (1997). Zero hour: Assessing the post-secondary system in Alberta. Calgary, AB: Alberta College and Technical Institutes Students' Executive Council.

AECD. (1993). Overview of department services. programs and challenges, 1992-93. Edmonton, AB: Government of Alberta, Ministry of Advanced Education and Career Development.

AECD. (1994). New directions for adult learning in Alberta. Edmonton, AB: Government of Alberta, Advanced Education and Career Development.

AECD. (1995). Institutional accountability in Alberta's post-secondary system: A progress report. Edmonton, AB: Government of Alberta, Advanced Education and Career Development.

AECD. (1996). Encouraging excellence and rewarding success in Alberta's public adult learning system: A proposal for implementing a performance funding envelope. Edmonton, AB: Government of Alberta, Advanced Education and Career Development.

AECD. (1997a). "$105 million for infrastructure renewal, Alberta universities, public colleges and technical institutes." News release. August 14. Edmonton, AB: Government of Alberta, Advanced Education and Career Development.

AECD. (1997b). "New performance funding rewards progress towards goals in post- secondary education." News release. July 31. Edmonton, AB: Government of Alberta, Advanced Education and Career Development.

AECD. (1997c). Rewarding progress towards goals: Highlights of Alberta's performance envelope funding. Edmonton, AB: Government of Alberta, Advanced Education and Career Development.

AECD. (1997d). Performance envelope press conference speaking points. Edmonton, AB: Government of Alberta, Advanced Education and Career Development.

AECD. (1997e). "Debt among Alberta students." September 2. Edmonton, AB: Government of Alberta, Advanced Education and Career Development.

Avery, B. (1997). "At NAIT, learning a trade means finding work." Edmonton Journal. September 25. p. A1, A20.

Barnetson, R.J. (1997). Marketing the University of Calgary to Frosh: A motivational typology of student-college choice. Master's Thesis. ERIC Document Reproduction Service. ED 404 930.

Bailey, A. (1991). "Library bloodletting." The University of Calgary Gauntlet. December 10. p. B5.

Calgary Herald. (1994). "Students default on $15 million." Calgary Herald. November 4. p. A1.

Cooper, D. & Neu, D. (1995). "The politics of debt and deficit in Alberta." in T. Harrison and G. Laxter. The Trojan horse: Alberta and the future of Canada. Montreal, PQ: Institute of Policy Alternatives of Montreal.

CAUT. (1996). "Average salaries for assistant professors." CAUT bulletin. March.

Drugge, S. (1995). The Alberta tax advantage: Myth and reality." in T. Harrison and G. Laxter. The Trojan horse: Alberta and the future of Canada. Montreal, PQ: Institute of Policy Alternatives of Montreal.

Emberley, P. (1996). Zero tolerance: Hot-button politics in Canada's universities. Toronto, ON: Penguin.

Gibbins, R. & Youngman, L. (1996). Mindscapes: Political ideologies towards the 21st century. Toronto, ON: McGraw-Hill Ryerson Company.

Gilbert, S. (1995). "Quality education: Does class-size matter?" Association of Universities and Colleges of Canada research brief. http://www.aucc.ca.

Government of Alberta. (1997). People and prosperity: A human resource strategy for Alberta. Edmonton, AB: Author.

Harvey, L. & Green, D. (1993). "Defining quality." Assessment and evaluation in higher education. 18. pp. 9-34.

Krahn, H. & Lowe, G.S. (1997). 1996 Alberta high-school graduate survey report. Edmonton, AB: Government of Alberta, Alberta Education.

Klein, R. (1996). An address at the Bomex '96 National Conference, September 5, 1996. http://www.gov.ab.ca .

Kohn, A. (1992). No contest: The case against competition. New York, NY: Houghton Mifflin Company.

Little, D. (1997). "Financing universities: Why are students paying more?" Education quarterly review. Statistics Canada. pp. 10-26.

March, J.G. (1976). "The technology of foolishness." in J.G. March and J.P. Olsen (Eds). Ambiguity and choice is organization. Oslo, Norway: Universiteforlaget. pp. 10- 23.

March, J.G. & Olsen, J.P. (1976). "Organizational choice under ambiguity." in J.G. March and J.P. Olsen (Eds). Ambiguity and choice is organization. Oslo, Norway: Universiteforlaget. pp. 54-68.

McMillian, M.L. & Warrack, A.A. (1995). "One-track (thinking) towards deficit reduction." in T. Harrison and G. Laxter. The Trojan horse: Alberta and the future of Canada. Montreal, PQ: Institute of Policy Alternatives of Montreal.

McQuaig, L. (1996). Shooting the hippo: Death by deficit and other Canadian myths. Toronto, ON: Penguin.

Robb, M. (1997). "U of A loses competitive salary edge: From second to sixteenth in a decade: Board chair suggests MLAs see data." University of Alberta folio. May 16. 34 (18). p. 1.

SAIT. (1996). Annual report. Calgary, AB: Southern Alberta Institute of Technology.

Stone, D. (1988). Policy paradox and political reason. Boston, IL: Scott, Foresman and Company.

Taft, K. (1997). Shredding the public interest: Ralph Klein and 25 years of one-party government. Edmonton, AB: The University of Alberta Press and the Parkland Institute.

Treasury. (1997). Budget update. Edmonton, AB: Government of Alberta, Treasury.

UCSU. (1996). Tuition, student aid and student debt: A comparison of Canadian and US universities. Calgary, AB: The University of Calgary Students' Union.

About the Author

Robert J. Barnetson
PhD candidate, Faculty of Education
The University of Calgary

Bob Barnetson is a doctoral student in adult education at The University of Calgary and Professional Affairs Officer at the Alberta College-Institutes Faculties Association.

Barnetson's most recent publication is "The Alberta Experience: A return to popular education?" in the *Canadian Journal for the Study of Adult Education*.

Mailing Address:
Alberta College-Institutes Faculties Association
#412 10357 109th St.
Edmonton, AB
T5J 1N3

voice (403) 423-4440
fax (403) 423-4515
email: barnetsonb@acifa.gmcc.ab.ca

Copyright 1997 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://olam.ed.asu.edu/epaa

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey@olam.ed.asu.edu .

EPAA Editorial Board

Michael W. Apple University of Wisconsin	Greg Camilli Rutgers University
John Covaleskie Northern Michigan University	Andrew Coulson a_coulson@msn.com
Alan Davis University of Colorado, Denver	Sherman Dorn University of South Florida
Mark E. Fetler California Commission on Teacher Credentialing	Richard Garlikov hmwkhelp@scott.net
Thomas F. Green Syracuse University	Alison I. Griffith York University
Arlen Gullickson Western Michigan University	Ernest R. House University of Colorado
Aimee Howley Marshall University	Craig B. Howley Appalachia Educational Laboratory
William Hunter University of Calgary	Richard M. Jaeger University of North Carolina--Greensboro
Daniel Kallós Umeå University	Benjamin Levin University of Manitoba
Thomas Mauhs-Pugh Rocky Mountain College	Dewayne Matthews Western Interstate Commission for Higher Education
William McInerney Purdue University	Mary P. McKeown Arizona Board of Regents
Les McLean University of Toronto	Susan Bobbitt Nolen University of Washington
Anne L. Pemberton apembert@pen.k12.va.us	Hugh G. Petrie SUNY Buffalo
Richard C. Richardson Arizona State University	Anthony G. Rud Jr. Purdue University
Dennis Sayers University of California at Davis	Jay D. Scribner University of Texas at Austin
Michael Scriven scriven@aol.com	Robert E. Stake University of Illinois--UC
Robert Stonehill U.S. Department of Education	Robert T. Stout Arizona State University